



## Factors influence data management model selections: IT Expert testimonies

### Dr. Gholam Ali Shaykhian, -

Gholam Shaykhian has received a Master of Science (M.S.) degree in Computer Systems from University of Central Florida and a second M.S. degree in Operations Research from the same university and has earned a Ph.D. in Operations Research from Florida Institute of Technology. His research interests include knowledge management, data mining, object-oriented methodologies, design patterns, software safety, genetic and optimization algorithms and data mining. Dr. Shaykhian is a professional member of the American Society for Engineering Education (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.

### Jinan Ziade ABD, University of Phoenix

My name is Jinan Ziade. Currently a Doctoral student, ABD, at University of Phoenix working on my doctorate of management in Organizational Leadership, specializing in Information System Technology. I worked for AT&T and had 14 years of experience in the Advertising and Publishing industry. During that time, I held many positions including senior graphic design, team leader, and project coordinator. Drawing on the many years of experiences and the varied background in the design and layout techniques I helped the company made decisions about allocating resources by providing the best foundation to analyze the workflow of knowledge and find solutions to improve employees' productivities and performance. I have also lead the United Way/AT&T campaign for the five consecutive years (2008, 2009, 2010, 2011, and 2012). In this role, I was responsible for developing new innovative ideas to raise funds and gained excellent supervisory and leadership skills. It was a great experience seeing all the work being done to ensure the United Way was spending the dollars collected wisely. Contributing time and effort was a great investment in the community. I enjoy traveling around the world with my son and my husband to broaden our awareness of other cultures, pick up a language or two, and meet interesting people. We have had some wonderful travels and enjoyed learning new cultures and discovered new excitements.

## **Factors influence data management model selection: IT Expert testimonies**

Gholam Ali Shaykhian, Ph.D.

Mohamed Abdelgadir Khairi, Ph.D.

Jinan Ziade, ABD

### **Abstract**

This paper examines the IT Expert testimonies perspectives to determine which factors (cost, schedule, performance, efficiency, limitations, risk, training, operations, compliances, deployment, security, accessibility, dependability, data quality, stability, maintainability, reliability, availability, flexibility, scalability, and predictability) influence data management model selections. Discussions encompass most widely used data management models namely Centralized Data Model (CDM) and Federated Data Model (FDM). A review of germinal and current literature reveals there are various aspects involved in data management such as the software development process, organizational culture, current IT trends, development methodologies, metrics, management techniques, terminology, leadership, learning organizations, quality, process improvement, and emotional intelligence. The study evaluates factors such as cost, schedule, performance, efficiency, limitations, risk, training, operations, compliances, security, accessibility, dependability, data quality, stability, maintainability, reliability, availability, flexibility, scalability, and predictability and how they influence CDM and FDM model selection and implementation. Each factor has its own unique influence on data management. The purpose of this paper is to address Expert testimonies as to which factors most influence data model selection and will cover all aspects and attributes that contribute to the architectural model selection.

### **Introduction**

Master Data Management, an IT discipline, is composed of a set of tools and processes to define master data entities of an organization. Master data such as customers, products, employees, materials, and suppliers are non-transactional key business data. MDM is essential to business operations and is the primary focus of this study. MDM objectives are to organize and manage the organization's master data. This paper discusses IT Expert testimonies on the two architectural models, Centralized and Federated, for MDM architectural implementation. In a Centralized Data Model (CDM), business consolidates master data in one repository [13]. Using CDM may resolve data duplications,

inconsistent master data, and improve data quality. However, implementing CDM may require users to overcome challenges such as geographical locations of the applications, cost of the implementation, and compliance with business rules and regulations.

Federated Data Model (FDM) allows an organization to extend data and business services to inquire data from multiple sources [12]. FDM's goal is to make master data available to all departments and partners of an organization. Yet, implementing FDM comes with many challenges such as synchronization of data between transactional and master data, network connectivity between the sources and MDM hub, performance, maintenance, and identifying roles and responsibilities.

This paper documents experts testimonies (15 authors/scientists who have published their research findings regarding the selection of CDM and/or FDM) on data management model selection; considering: cost, schedule, performance, efficiency, limitations, risk, training, operations, compliances, deployment, security, accessibility, dependability, data quality, stability, maintainability, reliability, availability, flexibility, scalability, and predictability. Experts are individual who have written books, journal articles and/or been involved in the design and architectural aspects of tools and technologies creations; they are the builders of the tool and have insight into the architectural frameworks of MDM.

## **Background**

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 research 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.

Master data consist of multiple entities. For example, master data may consist of customers, products, locations, and calendars [7]. MDM encompasses tools and processes to define and manage the non-transactional data (reference data that support company's transactional data, e.g. sales). The main objective of MDM is to offer tools for collection, aggregation, matching, consolidation, quality-assurance, persistence and distribution of master data all over an organization [3]. The use of master data in an unregulated, unplanned, and without architectural framework has been the source of data inconsistency and inefficiency. According to Kirby, Gile, and Fossner [8]. Inconsistent master data consumed by sales and marketing may negatively affect organization's performance.

Many IT organizations such as Microsoft have many databases spread across several divisions with no apparent consensus about specific data model architecture or shared MDM principles. Some divisions use FDM model and other divisions use CDM whereas the other divisions don't use MDM. This paper investigate the relevant attributes which

contribute in selection of the MDM models namely Centralized Data Model (CDM) and Federated Data Model (FDM) to help organizations to best bridge the gap among its divisions, technologies, and their stakeholders in MDM usage.

### **IT Expert Testimonies**

Twenty-One attributes, namely cost, schedule, performance, efficiency, limitations, risk, training, operations, compliances, security, accessibility, dependability, data quality, stability, deployment, maintainability, reliability, availability, flexibility, scalability, and predictability were considered as factors affecting the selection of CDM or FDM. The expert testimonies collectively agreed that all 21 factors influence on MDM model selection, however, no one specifically addressed them all as factors affecting model selection. Many of these factors are business requirements and technology driven factors that influence the MDM architectural model selection. The experts agreed on the fact that the factors such as cost, performance and deployment affect and influence the MDM model selection.

Experts such as Godinez, Hechler, Koenig, Lockwood, Oberhofer, and Schroeck [4], [15-16], if the cost of data consolidation and replication is greater than the business benefits, the FDM model should be used in MDM. In addition, to reduce cost in the beginning FDM should be used as a short-term practice [10] to drive quick benefits to the organization before moving to longer-term strategy (CDM) [4]. Shankar and Menon [14-15] argued the CDM is suitable with respect to cost. Lee stated that there are two main types of business cost that can be avoided by improved data management. First, there are the costs caused by poor data quality e.g., the need to verify data or poor decisions. Second, there are costs caused by improving or assuring data quality, such as the effort to prevent, detect, or repair poor data. Data quality cost could be reduced when implementing CDM.

CDM will benefit organizations for a longer-term strategy whereas FDM is useful for shorter-term strategy [4], [10]. The disadvantage of CDM is that the MDM system is not up to date like FDM system and that may delay the prompt delivery to current master data to consumers [4]. Operational MDM (FDM) can deliver significant gains in the form of operational efficiencies and process consistencies, but companies recognize the potential risk, costs and time to enterprise-breadth data to management of such a program and seek the executive sponsorship and funding required to make it a success [9]. Godinez [4], Shankar [13], Shankar and Menon [14], and Loshin [10] focused specifically on the factors influence have on FDM and CDM. They discussed how these factors influence CDM or FDM selection and the influence of business and technology on the master data management planning. For instance Shankar [13] explained the difference between business approach and technology approaches in model selection.

Godinez [4] believe FDM approach may not be a suitable to deal with large data because of the potential performance impact. However, FDM performance will improve if the MDM system queries small data from well-defined sources and was designed for short-term strategy. Friedman (2006) considered CDM with respect to performance because improved business capabilities are an important benefit of having a one source of a

company's data (CDM). Organizations are beginning to recognize that how they manage their key data is critical to their ability to improve agility and performance. Lack of trusted, quality information is widely seen as a strategic barrier inhibiting such capabilities as agility and competitive differentiation [16].

CDM host a single copy of master data that interact with MDM system whenever read or write to these master data records needed. MDM system becomes a source of authoritative truth of master data for an organization [4]. Operational MDM (FDM) can deliver significant gains in the form of operational efficiencies and process consistencies, but companies recognize the potential risk, costs and time to enterprise-breadth data to management of such a program and seek the executive sponsorship and funding required to make it a success [9]. Business problems are solved by MDM solution because it is managing master data that critical to the company [13]. Shankar [13] suggested that to have an efficient MDM solution based on focusing on the technology approach, organization should start simple, FDM model, and small with one entity, this way analytical usage will be easy, project risk reduces, data governance relieved, and the scope and ROI will be limited. However, the technology focus will not solve all business problems which lead companies to pursue CDM option [13].

FDM Compensate for SQL limitations at the data source by processing parts of a distributed request at the MDM federated server. Confounding factors such as legal constraints limit MDM across geopolitical boundaries [4], [10] do not allow CDM. Another limitation to use CDM is master data locking in Line of Business (LOB) packaged application [4].

A technology-focused approach begins with FDM (registry) model, one entity, and analytical usage. After the first step the MDM solution will become more mature to add more entities, different models such as CDM, and integrate the solution with operational systems. Shankar [13] found that starting with FDM approach allows organization to start quickly. However, this approach will be limited unless organization considers current business and possible future needs, this will allow organizations to get the right technology and have a better road map in place. Starting with CDM may be useful for the future growth and fast ROI [13].

The use of CDM provides complete master data view and enabling full risk assessment. CDM system allow assessing and understanding information assets risks such as operation risk (example denial of service), regularity and compliance risk, and reputational risk (for example customer data compromise lead to untrusted company) [4], [10]. All downstream applications depend on CDM system availability. The risk is high in case of such CDM system failure unless avoided by using redundant component in order to make MDM system highly available [4]. Operational MDM (FDM approach) can deliver significant gains in the form of operational efficiencies and process consistencies, but companies recognize the potential risk, costs and time to enterprise-breadth data to management of such a program and seek the executive sponsorship and funding required to make it a success [9]. Organization must consider the risk that the operational MDM (FDM approach) requires significant customization and custom coding of operational systems in order to enable the synchronization [9].

Training may not have direct influence in CDM or FDM models. However, knowledge gained by training adds needed skills to understand technology and tools and hence contributes on model's selection decisions. "It is important to realize that becoming a skilled practitioner of data modeling requires a significant amount of training and experience, so it is critical to engage individuals with the proper skill sets to successfully complete this task" [10]. "Training for enterprise roll-out of tools and technology make capabilities available on a more widespread basis" (p. 58). Loshin [10] stressed on the assessment of the training requirement to transfer from federated model to centralized model before starting transition in order to create smooth transition.

FDM implemented with two operational patterns: data integration and aggregation runtime; and availability and disaster recovery [4]. Godinez [4] added that in case of FDM the source systems operational is critical even if the MDM system is not. In the case of CDM high availability of variation of operational multi-tier for critical data pattern is needed. MDM services requests executed in MDM system using CDM in a transactional fashion and "thus part of the operational fabric of an IT environment" [4]. Operational risk such as Denial of Service (DoS) could be mitigated by using CDM which has centralized place to enforce security [4]. CDM is a dedicated system that stores a golden version of master data. In operational uses using the CDM approach, the master data published to downstream systems while source systems (upstream) maintain sovereignty to manage and own data. CDM Approach – consolidation approach – viewed in analytical use cases, clean and consistent master data published to data warehouse and BI applications [14]. In addition, Dagan [1] states that the CDM transcends operational and analytical use and into global governance on data management throughout the enterprise. Shankar and Menon [15] noted that having a compatible system between applications such as CRM and MDM system will lead to rapid operation.

FDM system could be a solution to the issue of not allowing the data to be copied out from its source due to security and licensing requirements. However, if source systems allow copying the master data to MDM system, CDM repository will provide secure environment for this data to be accessed by authorized users and applications. Enforcing security for MDM system is critical for business success and manages risks [4], [10]. Security in addition to performance, data governance, and high availability are most important aspects in implementing CDM because MDM in this model represents a single point of failure for applications accessing the master data. Improper security to a complete and up-to-date master data could be compromised by any attack [4]. Centralized security of master data is one of the advantages choosing CDM system. Loshin [10] suggested include access control as part of any governance structure in order to oversee the security management and identifying the access roles.

FDM system allows accessing data anywhere in the organization, without worrying about the format, vendor, new database creation, and disruptive changes to existing database using standard SQL or any tools supports Open DataBase Connectivity (ODBC) or Java DataBase Connectivity (JDBC) [4]. CDM as system of record provide centralized access, security, and governance. CDM can act as service consumer, service provider, or

both because it gets data from different sources that could be used by a requester [4]. Loshin [10] and Godinez [4] recommended on using FDM for short-term until the MDM is stable and gradually moves to CDM as a long-term strategy. FDM system is stable in a sense that data could be accessed from anywhere and anytime without creating new databases or without disruptive changes on existing ones, using standard SQL or any tools that support JDBC or ODBC.

Godinez [4] stated that MDM system deployed in FDM depends on availability of the source systems. The availability of the source systems decreases when it is numbers grows in FDM deployment. In contrary, all applications depend on availability of MDM system deployed in CDM. The advantage of CDM over FDM is that there is no CDM dependency on the source systems like FDM [4]. FDM is much preferable in the grid and cloud computing because the complexity of data management arises from the scale, vitality, undependability, and delivery of data sources [10]. One advantage of CDM implementation is that it is independent or not intrusive to the existing applications [4]. However, because CDM is a single point of enterprise-wide failure, all applications depend on its availability as oppose to FDM which depends on sources availability [4].

FDM may not be suitable approach when retrieving large data from the source with data quality problems [4]. In addition, source system's data that retrieved by FDM may not reach the quality bar that necessary for MDM system which lead to collaboration difficulties between the sources and MDM system [4] , [10]. The advantage of MDM system that implemented using CDM approach has ability to clean, valid, and matches the data to maintain high quality of master data [4]. Loshin[10] asserts that MDM deployed in CDM is responsible for the data quality because enterprise applications depends on to master data retrieved from MDM system and decisions made out of these data in contrary to CDM, FDM system is not responsible in data quality because the data quality is the sources' responsibility; this condition lower the cost of quality assurance. The focus of operational MDM (using FDM approach) is to make sure that data in multiple operational systems that should be the same is actually the same. The goal is to synchronize operational systems data so that you have consistency at the front end, such as in customer-facing systems. This goal is particularly important for organizations with a lot of customer contact.

FDM approach does not require sources data consolidation, therefore the maintenance of master data in FDM is minimal and it is up to the data source to maintain it [4]. Han and Chun [5] argue that maintainability of master data in CDM is not possible because master data needs it in multiple business systems and it could be maintained better in these systems. However, CDM application provides features to the users to maintain master data without data expert support attendance. Godinez [4] stated that companies have a fundamental problem in their ability to create reliable and accurate customer data for daily business processes. One major reason for data unreliability comes from incorrect and changeable customer reference data duplicated across the various IT systems within an enterprise. The unreliable sources impact the FDM approach negatively. A reliable master data management solution is to obtain master data from different sources and

consolidate data into one center (CDM approach): that accurate and consistent and could be used and consumed by all applications and users who needs it.

All systems enterprise-wide depends on availability of MDM (CDM Style) regarding master data because MDM in CDM style create a single point for master data to be consumed by all master data consumer in the enterprise applications. Contrary to FDM, CDM does not depend on availability of sources [4], [10]. High availability, performance, security, and data governance are the highest requirements for MDM implemented in CDM style [4]. CDM may help data warehouse vendors who are working to overcome scalability, replication, and availability concerns as systems grow into petabytes of storage, but have not yet provided solutions with the flexibility and benefits of an MDM SOA approach [13].

MDM (FDM style) is a flexible system because is allow flexible ways to ad hoc query master data from multiple sources [4]. The correlation of the data from remote data sources and local tables in FDM is seamless as all data stored in FDM database [4]. The flexibility of MDM (in FDM style) appears in ability of accessing data anywhere in the organization, regardless of data format or vendor, without new database creation, and affecting or disturbing existing databases, and using variety of access tools such as SQL, JDBC, and ODBC [4]. CDM may help data warehouse vendors who are working to overcome scalability, replication, and availability concerns as systems grow into petabytes of storage, but have not yet provided solutions with the flexibility and benefits of an MDM SOA approach [13].

With the interest on cloud computing growing, increases in grid applications in large scale of distribution and size, force organization to consider MDM in FDM style as a way to manage master data [4]. Dagon (2008) sees the urgent needs for scalable MDM system in CDM Style to handle complex business process, keep track of continual changes, and increasingly M&A activities. Scalable MDM system is able to transform systems to SaaS( software as a Service) by cultivating it to deal with multitenant environment as well as making sure scalability and configuration capability [5]. MDM implemented in CDM style provides predictable performance and high availability for master data consumption [4], [10]. If the number of sources feeds FDM with master data increases, the performance and availability becomes more unpredictable and that may impact the MDM solution [10].

The deployment of MDM in FDM style carry a risk as the number of sources grows because the availability can decrease quickly [4]. All applications in the enterprise depend on the availability of the MDM solution deployed in CDM style [4]. MDM solution deployed by FDM should provide business a metadata that provide the key linkages between critical data elements in the source systems [4], [10]. Deploying an MDM system in FDM style can be done quickly because source systems still manage most of the master data [4]. The MDM deployed in CDM environment allows the synchronization between an MDM system and other systems (master data consumers) unidirectional or bidirectional [4]. MDM is designed around the concept of a (virtual or physical) central repository (CDM) to store and manage master data and can be implemented according to various architectural styles [11].

## Conclusion

Expert testimonies testified how the 21 factors (cost, schedule, performance, efficiency, limitations, risk, training, operations, compliances, deployment, security, accessibility, dependability, data quality, stability, maintainability, reliability, availability, flexibility, scalability, and predictability) affect MDM's architectural model selection. Most experts concluded through their testimonies that CDM is the best option with respect to factors such as cost and availability. Some experts prefer CDM as an ultimate solution; they recommended FDM for short term approach. To reduce the cost at the beginning, FDM should be used as a short-term practice [10] to drive quick benefits to the organization before moving to longer-term strategy – CDM. Training factor is an example of factors that do not effect on model selection directly. Experts Testimonies shows the importance of training in MDM implementation in general.

Implementing the right MDM model will have a significant impact on the organization. The master data is one of the most sensitive and important assets for any organization. Management, protection, and accessibility of data are the core for IT success. Understanding the factors influencing the MDM models selection help leaders to plan properly and have complete awareness of the situation.

The results of this study should help leaders in their forecasting strategy. Understanding the 21 factors and their role on MDM implementation will allow them to forecast the cost, time, and data volume for implementation. If leaders understand the importance of the MDM models early and consider it in their strategy, then it makes for stronger leadership. Understanding the factors that affect MDM and potential problems that can be encountered in the different phases can give leadership the needful insight to prepare for potential problems and perhaps even avoid them.

## References

- [1] Dagan, B. (2008). Master data management systems to be useful for ccs, other functions. *Natural Gas*, 24(10), 25-29. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=31872515&site=ehost-live>
- [2] Danielewicz, J. (n.d.). SIX SIGMA AND MDM. *Data Strategy Journal*. Retrieved from [http://www.datastrategyjournal.com/index.php?option=com\\_content&task=view&id=64&Itemid=1](http://www.datastrategyjournal.com/index.php?option=com_content&task=view&id=64&Itemid=1)
- [3] Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M. (2008). Enterprise Master Data Management: An SOA Approach to Managing Core Information. IBM Press

- [4] Godinez, M. ; Hechler, E., Koenig, K., Lockwood, S., Oberhofer, M., Schroeck, M. (2010). The Art of Enterprise Information Architecture: A Systems-Based Approach for Unlocking Business Insight. Indianapolis, Indiana: IBM Press. Retrieved From <http://pqtechbus.safaribooksonline.com/book/databases/business-intelligence/9780137054947/enterprise-information-architecture-operational-model/ch06lev1sec1#searchlist>
- [5] Han, J., Ju, H., & Chun, J.(2010). RFID-ready Master Data Management for Reverse Logistics. Proceedings of World Academy of Science: Engineering & Technology, Jun2010, Vol. 66, p826-829, 4p
- [6] Juran, J. M. (1988). *Juran on Planning for Quality*. New York, NY: Macmillan.
- [7] Kimball, R. (2008). Essential Steps for the Integrated Enterprise Data Warehouse, Part 1. *swtuoproxy.museglobal.com*. Advance online publication. Retrieved from <http://swtuoproxy.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>
- [8] Kirby, P., Gile, C., & Fossner, L. (2006). Data warehouse architectures must reflect business consensus. *Forrester*. Retrieved from Microsoft Library
- [9] Longman, C. (2008). Why Master Data Management is Such a Challenge. *DM Review*, 18(11), 18-20
- [10] Loshin, D. (2008). *Master Data Management*. Morgan Kaufmann, CA: San Francisco
- [11] Lucas, A. (2010). TOWARDS CORPORATE DATA QUALITY MANAGEMENT. *Portuguese Journal of Management Studies*, 15 (2), 173-195
- [12] 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](http://www.infosysblogs.com/customer-relationship-management/2009/06/federated_mdm_data_domains_a_p_1.html)
- [13] 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/>.
- [14] Shankar, R. (2008). Master Data Management Strategies to Start Small and Grow Big. *Business Intelligence Journal*, 13 (3), p37-47, 11p

- [15] Shankar, R., & Menon, R. (2010). MDM Maturity Pragmatism Business Challenges and the Future of MDM. BUSINESS INTELLIGENCE Journal. 15(3) p19-25
- [16] White, A. and Y. Genovese. (2006). "Your Business Process Platform Needs and Enterprise Information Strategy," GartnerGroup Research, April 26, ID# G00139332.