

---

## Master Data Management: Emerging Issues

<sup>1</sup>**Praveen Kumar Gupta,**

Professor, Jagan Institute of Management Studies, Guru Gobind Singh Indraprastha University, Delhi.

<sup>2</sup>**J. P. Singh,**

Professor, Jagan Institute of Management Studies, Guru Gobind Singh Indraprastha University, Delhi.

<sup>3</sup>**Jagdeep Kaliraman,**

Assistant Professor, Aravali College of Engineering and Management, M. D. University, Haryana.

*Abstract: The objective of this paper is to make a concerted effort towards investigative analysis and eliciting the challenges belonging to master data management (MDM) domain in the area of data science. This is a review work. The paper begins with data science and focuses on master data management and considers different parameters and comparing the key concepts of various aspects of master data management. Finally, in the context of data science, the paper discusses with the potential emerging issues related to master data management with a view to help in designing an effective and ethical MDM.*

**Index Terms-** Master Data Management, Databases, Data-warehouse, MDA, normalization.

### I. INTRODUCTION

The current interest in the data science domain is largely due to the volume, velocity and variety of data generated especially in the second half of the 20<sup>th</sup> century. The structured and unstructured data handling has resulted into the pain that organizations are experiencing around consistent reporting, regulatory compliance, strong interest in Service-Oriented Architecture (SOA), and Software as a Service (SaaS) has prompted a great deal of interest in Master Data Management (MDM). This paper explains what MDM is, why it is important, and how to manage it, while identifying some of the key MDM management issues that are emerging. This paper evolves from MDM to the technical and procedural emerging issues involved in MDM.

### II. REVIEW LITERATURE

MDM is fundamental to managing data. The software systems have basic data that are shared and used by multiple applications that constitute the system. Say, an enterprise resource planning (ERP) system as a minimum, will have masters of a Customer, an Item, and an Account. This master data is the key asset of the company. Many acquisitions in corporate world take place primarily for access to its Customer Master data. There are some very well-understood and easily identified master-data items, such as "customer" and "product." In fact, many define master data by simply reciting a commonly agreed upon master-data item list, such as: customer, product, location, employee, and asset. But how you identify elements of data that should be managed by a master-data management system is much more complex and defies such basic definitions. In fact, there is a lot of confusion around what master data is and how it is qualified, requiring a more comprehensive treatment.

There are essentially five types of data in corporations:

- Unstructured—This is data found in e-mail, magazine articles, corporate intranet portals, product specifications, marketing collateral, and PDF files.
- Transactional—This is data related to sales, deliveries, invoices, trouble tickets, claims, and other monetary and non-monetary interactions.

- **Metadata**—This is data about other data and may reside in a formal repository or in various other forms such as XML documents, report definitions, column descriptions in a database, log files, connections, and configuration files.
- **Hierarchical**—Hierarchical data stores the relationships between other data. It may be stored as part of an accounting system or separately as descriptions of real-world relationships, such as company organizational structures or product lines. Hierarchical data is sometimes considered a super MDM domain, because it is critical to understanding and sometimes discovering the relationships between master data.
- **Master**—Master data are the critical nouns of a business and fall generally into four groupings: people, things, places, and concepts. Further categorizations within those groupings are called subject areas, domain areas, or entity types. For example, within people, there are customer, employee, and salesperson. Within things, there are product, part, store, and asset. Within concepts, there are things like contract, warrantee, and licenses. Finally, within places, there are office locations and geographic divisions. Some of these domain areas may be further divided. Customer may be further segmented, based on incentives and history. A company may have normal customers, as well as premiere and executive customers. Product may be further segmented by sector and industry. The requirements, life cycle, and CRUD cycle for a product in the Consumer Packaged Goods (CPG) sector is likely very different from those of the clothing industry. The granularity of domains is essentially determined by the magnitude of differences between the attributes of the entities within them.

Next issue is on deciding on what constitutes MDM. While identifying master data entities is relatively easy, not all data that fits the definition for master data should necessarily be managed as such. We can follow criteria for deciding master data as discussed below, all of which should be considered together when deciding if a given entity should be treated as master data.

Master data can be described by the way that it is created, read, updated, deleted, and searched. This life cycle is called the CRUD cycle and is different for different master-data element types and companies. For example, how a customer is created depends largely upon a company's business rules, industry segment, and data systems. The following table illustrates the differing CRUD cycles for four common master-data subject areas.

### Sample CRUD cycle

	Customer	Product	Asset	Employee
Create	Customer visit, such as Web site or account created	Product purchased or manufactured; involvement	Unit acquired by HR hires, numerous forms, opening a PO; orientation, approval process selection, asset allocations, necessary office assignments	by HR hires, numerous forms, PO; orientation, benefits allocations, office assignments
Read	Contextualized views based on credentials of viewer	Periodic inventory catalogues	Periodic reporting purposes, depreciation, verification	reporting figuring Office access, reviews, insurance-claims, immigration

---

Update	Address, phone preferences, accounts	discounts, number, credit raw materials	Packaging changes, changes	Transfers, maintenance, reports	Immigration status, marriage status, level increase, raises, transfers
Destroy	Death, liquidation, do-not-call.	bankruptcy, Canceled, no longer available	replaced,	Obsolete, destroyed, scrapped	sold, stolen, Termination, death
Search	CRM system, call-center system, contact-management system	ERP system, orders-processing system	GL tracking, asset DB management	HR LOB system	

Finally, when deciding on master data, it is better to categorize entities in terms of their behavior and attributes within the context of the business needs than to rely on simple lists of entity types.

Master Data Management (MDM) includes the technology, tools, and processes required to create and maintain consistent and accurate lists of master data. MDM requires fundamental changes to business process to maintain clean master data, and some of the most difficult MDM issues are more political than technical. Another thing to note is that MDM includes both creating and maintaining master data. Investing a lot of time, money, and effort in creating a clean, consistent set of master data is a wasted effort unless the solution includes tools and processes to keep the master data clean and consistent as it is updated and expanded.

The two main categories of tools are Customer Data Integration (CDI) tools for creating the customer master and Product Information Management (PIM) tools for creating the product master. Some tools will do both, but generally they are better at one or the other.

MDM is a complex process that can go on for a long time. Like most things in software, the key to success is to implement MDM incrementally, so that the business realizes a series of short-term benefits while the complete project is a long-term process. No MDM project can be successful without the support and participation of the business users. IT professionals do not have the domain knowledge to create and maintain high-quality master data. Any MDM project that does not include changes to the processes that create, maintain, and validate master data is likely to fail.

### III. CREATING A MASTER LIST

There are two basic steps to creating master data: cleaning and standardizing the data, and unify data from all the sources to remove duplicates. Before the first step, we must understand the data model for the master data. The modeling process includes the contents of each attribute defined, and a mapping needs to be defined from each source system to the master-data model. This information is used to define the transformations necessary to clean your source data. Cleaning the data and transforming it into the master data model is very similar to the Extract, Transform, and Load (ETL) processes used to populate a data warehouse.

---

Matching master-data records to eliminate duplicates is both the hardest and most important step in creating master data. The matching accuracy of MDM tools is one of the most important purchase criteria. So, matching algorithms are normally very complex and sophisticated.

Another factor to consider when merging your source data into the master list is privacy. When customers become part of the customer master, their information might be visible to any of the applications that have access to the customer master. If the customer data was obtained under a privacy policy that limited its use to a particular application, you might not be able to merge it into the customer master. So, one may add a lawyer specializing in intellectual property rights (IPR) to your MDM planning team.

#### IV. MDM EMERGING ISSUES

One of the leading challenges is synchronization leading to data inconsistency or out of date, resulting into flawed decision based on flawed data.

Another challenge is handling the temporal data and the market's uncertainty, may make the selection of technology components difficult, which may result into committing to a failing product/technology.

Another interesting challenge is Integration of business process management (BPM) with MDM since MDM fundamentally changes the business processes and this will require the stakeholder's involvement and acceptance, an important part of organisation's culture.

On-premises MDM has the inherent challenges as mentioned above.

Cloud has added another dimension to MDM. On one side, MDM in cloud provides significant cost savings, scalability, flexibility and improved accessibility, there are associated potential challenges like security, reliability & loss of control on the organizational data and apps. The research firms like Gartner and Magic Quadrant have excluded cloud based MDM services from their research report because cloud based MDM services are so new and have yet to affect the on-premises MDM deployments.

MDM is fundamental to Data Science. The application of data science in data analytics requires handling less of structured and more of unstructured data. The technology is sufficiently mature to handle structured data but it is still evolving w.r.t. the unstructured data. So, analyzing usefully the unstructured data is a challenge. One such environment and tool that helps us in this area is R programming and environment. R is a programming language and environment for statistical computing and graphics, suitable for data analytics. Because of its powerful built-in tools and large library for R programming, it makes R very flexible. R is very practical and is meant for actual use and not just academic research). Since R belongs to open source community, co-reuse and value addition is the core philosophy, without reinventing the wheel. R is one of the well accepted industry standards. Skype and Opera(browser) have been created using R is a testimony to this.

The ethical issues are of great importance in contemporary Data Science Research. It must be ethical in the local context in this global village. A successful MDM based Data Science research must provide stability and flexibility. Without stability, ethical pre-commitments will vanish as technology rapidly shifts costs.

---

## V. CONCLUSION

The recent emphasis on regulatory compliance, SOA, and mergers and acquisitions has made the creating and maintaining of accurate and complete master data a business imperative. Both large and small businesses must develop data-maintenance and governance processes and procedures, to obtain and maintain accurate master data. While it's easy to think of master-data management as a technological issue, a purely technological solution requires corresponding changes to business processes and controls to succeed and produce satisfactory results. This paper has explained master-data management, the process of evolving a solution, and technological implementation along with the emerging issues.

## REFERENCES

- [1] Wainwright, Phil, "How Master Data Management Supercharges Fast Moving Businesses", White Paper published by Procullux Corporation, 2013.
- [2] Kaliraman, J. et. al, "Vehicle Security System and Accident Information System by using VLSI", conference proceedings of ICICCT, Volume 1, Excel Publishers India, pp 122-130, 2016.
- [3] Gupta Praveen Kumar, "Fuzzy Functional Dependencies in Temporal Databases", 2016, PhD thesis, Computer Science and Engineering Department, UIET, MDU, India.
- [4] Gupta Praveen K, Rishi, R. et. al, "An investigative analysis of timestamping approaches of temporal models", conference proceedings ICACCT, publisher – Inderscience, UK, 2013.
- [5] Gupta, Praveen k., "An approach towards seamless integration of Enterprise resource Planning (ERP) with fuzzy perspective", research paper published, International Journal of Fuzzy Systems and Rough Systems, Volume 1(1), June 2008, page 31-36.
- [6] Gupta, Praveen kumar, "Intellectual Property rights: emerging issues", Paper contributed at National Seminar on Artificial Intelligence And soft computing, ITM, Gurgaon, 29-30 August 2008.
- [7] Gupta, Praveen kumar, "An investigative analysis of challenges in grid computing", research Paper published with the national journal of Computer Science and Information Technology, ISSN: 0973 – 4872, June 2008
- [8] Loshin, David, "Addressing Emerging Challenges in Big Data", White Paper published with www.progress.com, Progress Software Corporation, 2016.
- [9] Wolter, Roger et. al, "The What, Why and How of Master Data Management", White Paper published by Microsoft Corporation (msdn.microsoft.com), 2006.
- [10] Gupta Praveen K., Biswas, R. et. al, "A comparative analysis of temporal data models", IJACEN, 2013.