

White Paper



Entity Standardization

An MDM Challenge

Jairaj Asok Kumar

Abstract

Master Data management, an enterprise-wide initiative, discuss about having a centralized repository of all enterprise entities, be it customer, product, account or location. The centralized hub facilitates a 360o degree view of the enterprise entities and truly represents the entities including their relationship, dependencies, audit and history over time, etc. However over period of time the “master data” undergoes data decays and all Master Data Management implementation depends on Entity Standardization. So what is Entity Standardization? This means “Golden copy” of the master data is achieved through internal procedures defined as rules/ criteria or through integration with external tools or even third party CDI integrators. Entity standardization discusses in details the mode of the managing the trueness of master data in terms of recency and correctness. This paper delves into a typical case study where an internal Entity Standardization tool is used for Customer entity standardization in terms of Name, Address, Organization, Phone and Person Matching (suspect determination) etc.

Entity Standardization – an MDM implementation Challenge

Master data Management (MDM) is about managing “master data” from diverse systems, having diverse representation, into a centralized hub - the MDM solution. Master Data management is therefore a set of process, procedures and standards by which master data or “Enterprise wide Entities”, herein referred to as the multi-form, multi-domain representation, is managed as Information across the enterprise. Information, the life blood of any organization, is derived from underlying data and is transient with passage of time. E.g. Mary Jane may get married and her new name would be Mary Roberts, but irrespective of the difference in “maiden name” attribute refers to the same person. Another illustration is Joe Harry residing in 101 Poplar Street, NJ0237 moving to a new address as part of job transfer to a revised address of 201 North Tryon Street, NC2310. The data capture points for each of these instances may be manual (with errors occurring during the data capture process) or process may not tag the data set with the correct entity or may not be captured at all. Each of the above illustration leads to data decay of the “golden copy” and how do you manage data decay and ensure that the “golden copy” or non-redundant 360 view of customer is managed over time?

Master Data management solutions (be it product domain or customer domain) are enablers towards focused customer centric initiatives. To achieve the customer centric objectives it is required to obtain a single representation of the customer (golden record) along with a 360 degree view of the customer. It is further complicated when there exists multiple Line of business having diverse system, each with a different representing of customer and customer specific attributes. How are the problem related to “golden copy” and “unique representation of master data” resolved? These would be addressed in the next section.

Why is Entity Standardization a key challenge for Master Data Management Implementations?

Typical MDM implementation requires migration of Master data from diverse system into a centralized data hub. During this process the business nomenclature of the key entities vary in terms of structure, domain value compliance and form. It is then typically to recommend Entity standardization software, be it through an internal tool (e.g. Trillium, QualityStage or in-built rules engine) or through an external Customer Data Integrator partner (e.g. Duns and Bradstreet or Acxiom). “Entity Standardization” can thus be defined as the process of defining a “Golden copy” of the master data (Customer, Account, Location and Product) which is achieved through internal procedures defined as rules/criteria/semantics or through integration with external tools or even third party Customer Data Integrator (CDI) integrators. Entity standardization discusses in details the mode of the managing the trueness of master data in terms of recency and correctness.

Now, that we have defined “Entity Standardization” what does it constitute? At a minimalistic level it includes the following

1. Person name standardization – typical for customer entity.
2. Address and phone standardization – typical for customer demographics.
3. Product data mastering- typical for customer product holding.
4. Customer and Organization standardization –typical through pattern based or match/merge data rule set.
5. Product standardization – typical through semantic transformation.

The diagram depicted below shows a quick representation of the various cogs in a typical “entity standardization” scenario.

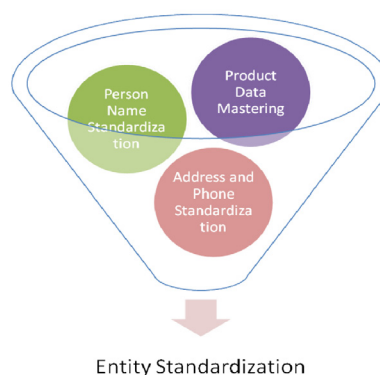


Figure 1 - Entity Standardization – Overview

As an organization undertaking the Entity Standardization process, it is essential to leverage its existing entity standardization toolset if available (be this QualityStage, Trillium, SilverCreek, Dataflux etc). If the quality of the master data is really bad, with multiple data representation and limited to latent demographic representation, it would be ideal to use third party CDI integrators (be this Acxiom, Duns and Bradstreet, FirstLogic or Experian). Each of the above entity standardization technique has its advantages and disadvantages. A comparison of these factors head to head for each of these techniques is depicted below.

Factors	Tool based Entity Standardization	Third party External Entity Standardization
Quality of data	Master data quality is assessed to be relatively high	Master data quality is assessed to be relatively low
Compliance and regulatory risk of data exposure	Master data has been classified as level 1 information (classified)	Master data has been classified as level 3 information (sensitive)
IT landscape	Diverse tools and data bases with existing Entity standardization tool	Limited data store and no existing entity standardization tool.
Business intelligence awareness	Highly aware	Limited awareness
Customer unique identifier	Plan to define a global identifier only at enterprise level	Plan to define a global identifier across supply chain.
CASS support	Required and needs integration with multiple consuming applications	Not available and limited support for multiple consuming applications
Cost factor	Initial upfront - capital cost	Cheaper upfront cost

Figure 2 - Entity Standardization Techniques - Head/Head comparison

Business drivers to take up Entity Standardization process

The key business drivers to take up Entity Standardization process is listed below.

1. Empowering risk and compliance initiatives with the information they require. – Ensuring that customer address preference is captured along with a do-not-call registry.
2. Optimizing revenue opportunities by ensuring effective and efficient interactions with customers, partners, and suppliers. – Ensuring product and customer standardization i.e. name, address or product holding data.
3. Enabling collaborative business processes with consistent and trustworthy information. – Ensuring common Meta data is shared across the various stakeholders.
4. Reducing the total cost of ownership for maintaining consistent information across the enterprise. – Enterprise wide representation of Customer and product data.

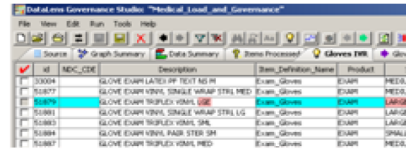
Entity standardization the maturity curve

Entity Standardization is not a buzz word. Any project be it data migration, application rewrite, data warehousing or data management initiatives (Meta-data or master-data), has rudimentary entity standardization process built in. So what would be the maturity curve of entity standardization tool? The entity standardization tool typically uses the following

- Coded rules
- Visual rules
- Auto learn/Interfaces

The first generation Entity Standardization tools, used Coded rules i.e. basically syntactic (based on patterns) and are hard-wired into the transformation logic. They are put together by the IT department and hence required manual intervention and are inflexible. The next generation entity standardization tool used visual rules that were based on semantics and were modified by the SME. The Meta data rules were then incorporated during the normal CRUD (Create, Read, Update or Delete) transaction where the entity were persisted, stored or retrieved. The latest 3rd generation Entity standardization tools, uses auto learn/interfaces that are auto assisted by the system when the non-technical SME's defines the semantics. These tools have good data effectiveness and efficiency in providing “data lineage” i.e. understand what process has been used to change the master data during the entity standardization process.

3rd Generation:
**AutoLearn/
Inference**



ID	MDX_CODE	Description	Exam_Definition_Name	Product	IS
51004		GLOVE EXAM LATED RP TEXT NS IN	Exam_Gloves	EXAM	MEDLP
51007		GLOVE EXAM VINYL SINGLE WRAP STRL MED	Exam_Gloves	EXAM	PRESLP
51006		GLOVE EXAM TRIPLEX VINYL GLO	Exam_Gloves	EXAM	LARGE
51001		GLOVE EXAM VINYL SINGLE WRAP STRL LG	Exam_Gloves	EXAM	LARGE
51003		GLOVE EXAM TRIPLEX VINYL SM	Exam_Gloves	EXAM	LARGE
51004		GLOVE EXAM VINYL PWR STRL SM	Exam_Gloves	EXAM	SMALL
51007		GLOVE EXAM TRIPLEX VINYL MED	Exam_Gloves	EXAM	MEDLP

Built by: System – assisted by non-technical SME
Method: **Semantic** (based on context)
Reuse: Very good (generalizes well)
Timeline: Hours

2nd Generation:
Visual rules



Built by: SME
Method: **Semantic** (based on context)
Reuse: Very good (generalizes well)
Timeline: Months

1st Generation:
Coded rules



Built by: IT
Method: **Syntactic** (based on patterns)
Reuse: Poor (does not generalize)
Timeline: Years

Figure 3 - Entity Standardization - Maturity View

Best Practices to overcome Entity Standardization challenges

Now that we have delved into what constitutes entity standardization, let us have a quick look at the best practices to be used for overcoming Entity Standardization Any initiatives addressing Entity Standardization requires the identification of best practices and this section aims to quote a few instances.

- Embark on a fully fledged Data profiling exercise upfront.
- Agree upfront on the process for resolving data standardization issues that are ambiguous.
- Defining semantics that can adapt to the entity representation as they undergo standardization.
- Developing integrated data governance, where data visibility and process visibility of entity standardization can be monitored and managed.
- Involving sense of ownership by the Data owners or business stakeholders.
- Embark on an external third party CDI integrator if the quality of data is really bad. The advantages that a third party CDI integrator brings are higher quality data with support for data augmentation and unique identifier generation.

CASE STUDY

Entity Standardization

This section talks about a real-time scenario where the largest retailer based out of Australia aims to diversify into retail banking, by launching a credit card. Typically a retailer has its core operational process (inward focusing) that is efficient, but when it comes to customer centricity there is always a lack in the data quality of the master data. To overcome this data-fatigue, the client embarked on a fully fledged customer centric initiative, driven through implementation of Master Data management and Service orientation. Understanding the poor quality of customer data, it embarked on Entity standardization by using a tool based asset (QualityStage). QualityStage (or be it any other tool), has default entity standardization for Person Name, Organization Name, Phone number. Additionally it offers a fully-fledged match/merge capability defined using Meta data (rules based assessment).

The master data capture point and master data consumption point, are typically multi channel and multi system, requires a uniform representation of the master data and usually achieved during the migration of the master data from the internal data sources. Further on every transactional operation on the master data based on trust levels, update were done – i.e. where the master data is obtained from the financial partner, it was considered to be having a higher value of trust. This reduces the number of manual intervention, for performing Entity Standardization. By using this underlying principle the internal master data or key customer information was refreshed, to have the best possible customer demographics representation. Additionally QualityStage was integrated as part of real-time transaction to support Person Name, Address and Phone standardization. The Out Of the Box (OOTB), capability of rule based data match and merge where suspects are promoted to either as suspect or new customer was implemented in QualityStage. The usage of the Entity Standardization tool helped shorten the time of implementation of a quick “win-win” MDM deployment.

Conclusion

Entity Standardization is an essential element during an MDM implementation. The offshoot of Entity Standardization programme is the ability of an organization to achieve optimal ROI for its customer centric strategies and to ensure a uniform customer experience management across multi-channel by its loyal customers. Information understanding is an increasingly important organizational issue. Most critical business initiatives depend of quality information. Improving understanding requires a focused approach including business, data, and system levels. Effective communication delivers results across these boundaries. At the core of any Information Quality initiative is a platform capable of providing ongoing understanding of the master Data entities. Entity Standardization helps in achieving the “information driven objectives” of an organization.

References

1. Data Quality Discovery and Analysis Best Practices with Information Analyzer - Harshita Mehta, IBM InfoSphere Consultant, IBM
2. Silver Creek system – an overview

About Author

Jairaj is a Lead consultant at the Master Data Management (MDM) practice in Infosys. He has more than a decade of experience in leading the delivery of information management solutions (web application and business intelligence applications) for global companies in banking, insurance, high-end technology and retail in the U.S., U.K. and Australia.

Jairaj is an expert in MDM service-based offerings in data rationalization and consolidation, skill building through training and developing assets, re-usable artifacts and templates for MDM. He also has experience in information management consulting, channel partner accreditation/business process re-engineering/business transformation using web application and business intelligence/data warehousing.



For more information, contact askus@infosys.com

About Infosys

Many of the world's most successful organizations rely on Infosys to deliver measurable business value. Infosys provides business consulting, technology, engineering and outsourcing services to help clients in over 30 countries build tomorrow's enterprise.

For more information about Infosys (NASDAQ:INFY), visit www.infosys.com.