

Evolutionary Multi-Domain MDM and Governance in an Oracle Ecosystem

FX Nicolas
Semarchy

Keywords:

Master Data Management, MDM, Data Governance, Data Integration

Introduction

Enterprise ecosystems have grown (organically or by acquisitions) to include many components of the Oracle stack, from both the Middleware and Applications side. This natural development frequently leads to an anarchic distribution of the critical data assets (customers, products, financial hierarchies, reference data, etc.) across applications and systems. Multi-domain master data management aims at structuring, governing, managing and linking this critical data for better operational efficiency, decision making and regulatory compliance.

What is Master Data?

Before discussing Master Data Management, we must first define master data. The following classification of data is commonly agreed in the data management field. We have seen it in the past in the form of a stack, a pyramid, or even a diamond as represented below; regardless of the shape, the list of items remains the same.

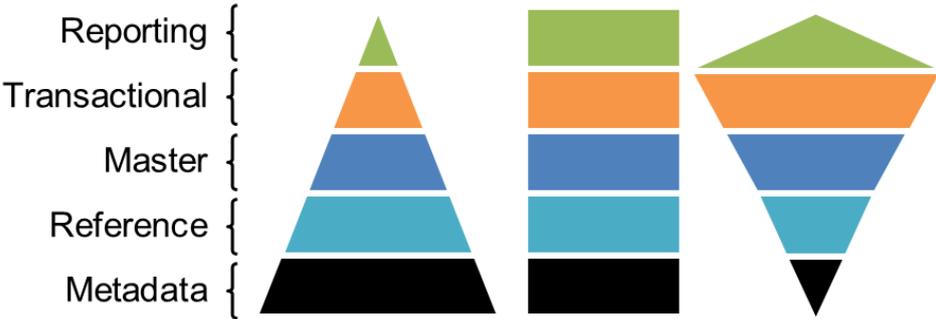


Figure 1: Data Classification

Master Data/Reference is key business information that supports the transactions. Master Data describes the customers, products, parts, employees, materials, suppliers, sites, etc. involved in the transactions. It is commonly referred to as Places (locations, geography, sites, etc.), Parties (persons, customers, suppliers, employees, etc.) and Things (products, items, material, vehicles, etc.). Reference data describes the standards (ISO, EC1@ss, etc) on which all data rely.

Master data already exists and is used in the operational systems, with some issues. Master data in these systems is:

- Not high quality data,
- Scattered and duplicated,
- Not truly managed.

Master Data is usually authored and used in the normal course of operations by existing business processes. Unfortunately, these operational business processes are tailored for an “application-specific” usage of this master data and therefore fail in achieving the overall enterprise requirement that mandates a common set of master data for applications with high quality standards and common governance.

The Challenge of Master Data

Applications used by each division/department within a company are frequently best of breed applications, each storing its version of certain master data with no consideration for the other application, and sometimes with no consideration for the quality of this master data.

Other challenges include the applications market consolidation (Oracle is a preeminent company here), and the mergers and acquisition exercise to which most companies are sooner or later confronted to. A direct consequence of these two challenges is that most companies do not run with one ERP or CRM flavor, but with multiple flavors of those applications. Add to the picture the fact that even within a single flavor of application, there are frequently multiple instances. It is obvious that key data assets like "customers" and "products" are likely to be scattered, low in quality and impossible to manage only with the applications in place.

Transactional and reporting data heavily rely on master (and reference) data. As a consequence, “bad master data” reflects directly into untrustworthy reports and operational inefficiency.

Master Data Management offers the possibility to take control over critical data with a single point of truth where this data is standardized in form and content, made complete and unique. This high quality standardized data simplifies data exchange and integration with applications, business partners and facilitates mergers and acquisitions as new systems join a well-organized set of foundational data.

The MDM Hub Patterns

There are various patterns or styles identified in the industry for implementing such MDM hubs and creating golden data.

- **Registry Hub Style:** In this style of hub, master data is authored and remains in the source systems. The hub stores an index of this source data, keeping track of the cross-references between matching source data. This hub style also typically stores the attributes used for matching purposes.
- **Consolidation Hub Style:** In this pattern, master data is copied from source applications, then matched and consolidated in the hub; the golden data physically created in the hub becomes available for distribution to downstream applications or for direct consumption by business users and data stewards.

- Co-Existence Hub Style: This pattern is similar to the Consolidation hub. It adds an integration loop back to the source applications for Golden Data.
- Centralized/Transactional Hub: In this pattern, master data is entirely migrated and cleansed in the MDM hub. The hub becomes the single provider for master data, and master data is exclusively authored in the hub through data entry workflows. All applications refer to the hub for their master data.

None of the styles described above is a magic pattern (or golden ring) to rule all the master data.

Depending on the domain, the source applications, or the master data that you plan to have in your MDM hub, you may want to store all the data or only the references, or something in between. You may want to enable consolidation only, with some data entry capabilities, or data entry only for certain data.

The choice of a style is not a technical decision. It must be driven by the business requirements. In addition, this choice is not written into stone. It evolves over time. Hence, the hub should support all these styles out of the box.

Why Multi-Domain MDM?

Master Data Management platforms originally were designed to fix the lack of proper master data management for two types of entities: the Customer and the Product. Customer Data Hubs (CDH) were sold and attached to CRM platforms to detect and handle customer duplicates instead of the CRM. At the same time Product Information Management (PIM) solutions were attached to the ERP or SCM products to handle product data in a highly collaborative way.

At a certain stage, other master data emerged and required their own tools for management: financial hierarchies, employees, suppliers, etc. Some could fit in a CDH; some could fit in a PIM; some called for their own tools (Oracle DRM – focused on hierarchies – is an excellent example).

Yet, fitting a customer in a PIM or a product in a CDH was not (and is still not, despite all vendors' efforts to do so) really possible, as it would have been the equivalent of using a CRM to manage Products or a PLM to manage Accounts and Contacts. A very long stretch.

After a few years of market evolution, Multi-domain MDM, which is the capability to manage and relate any type of master data in a single platform, emerged.

Metadata & Logical Design

Looking at the requirements and hub styles for MDM, one may argue that a data quality tool, a data integration product and a database, one can build an MDM hub from scratch.

To expand on this idea in a pure Oracle ecosystem:

- The Oracle database would provide the storage and access capabilities for the MDM hub, and the processing power.
- Oracle Data Integrator, or components from SOA Suite would provide the movement and transformation capabilities within the hub.
- Oracle DQ tools would provide the advanced data quality, cleansing and de-duplication features.

- UIs, Business processes and workflows would be managed with a mix of Oracle BPM, Human Workflow and ADF.

Even with a team of skilled developers in all these technologies, one would still need to manage metadata across the board.

With all this information, it should be possible to automatically generate the bits and bytes of the MDM hub, including:

1. Physical structures for storing the golden data.
2. Endpoints (inputs) for applications to publish their raw data to the hub
3. Endpoints (outputs) for applications to consume the master data from the hub
4. Processes and structures to automatically enrich, validate, match, de-duplicate and certify golden data from the raw (source) master data.
5. User interfaces to view and edit the master data as well as the various stages of the process described above.
6. Human workflows to validate and correct erroneous data

This is the “magic” that you should expect from a true MDM platform. With a completely metadata-driven approach, you would focus on what really counts: *the metadata, which is in essence the knowledge of your master data.*

The success of data integration (ELT) platforms such as Oracle Data Integrator is similar at a technical level. By aggregating common integration and data cleansing technical patterns known to every data-warehousing practitioner in the form of code templates, this platform re-focused the effort of data integration to its real objective, which was expressing the integration rules (mappings) and just applying the patterns. Endless re-coding (understand: copy-pasting) of well-known patterns was history. If the same applies to MDM, we need to define the objective of MDM: It is to quickly implement the decisions taken by the Data Governance processes.

Evolutionary MDM in an Oracle Ecosystem

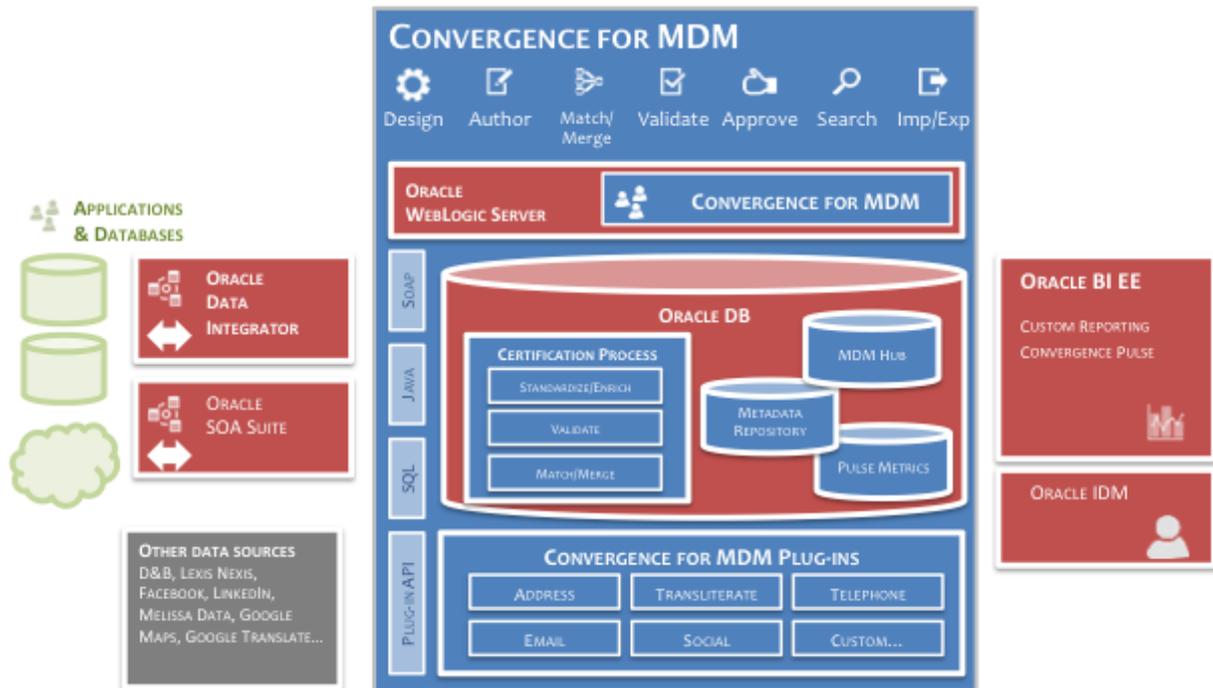
What is defined as Evolutionary MDM is MDM that:

- Does not mandate a specific hub style but instead adjusts the hub style to the governance requirements and to the corporate ecosystem
- Provides the level of agility, flexibility and scalability to support any domain and the pace of changes required by data governance.
- Support data measurability, security, compliance, lineage, and traceability as requested by data governance.

It is, in a nutshell, MDM that evolves with the business.

MDM Blueprint

The following blueprint describes an evolutionary MDM architecture in an Oracle ecosystem, using the Semarchy Convergence for MDM platform.



Blueprint Components:

- The Convergence for MDM JEE application is deployed in an Oracle WebLogic Server.
- Oracle Identity Manager provides authentication and Roles to the application through WLS.
- Oracle Database 12c (All editions supported) is used as the backend for storing the metadata, the master data hub as well as the hub metrics (Pulse Metrics).
- The master data processes run within the Oracle Database. These include (among others):
 - Searching
 - Enriching/Standardizing
 - Validating Data Quality rules
 - Matching and merging duplicates
- External data sources are invoked using built-in Java plug-ins, which can be extended by customized ones.
- Frontend for runtime and design-time user interfaces are rendered in a web browser.
- Oracle Data Integrator and/or SOA Suite handle integration with applications and databases through Tables, Web Services and Java APIs generated by Convergence for MDM.
- Profiling and hub metrics are captured by Convergence for MDM into dedicated data stores that are used by Convergence for MDM's built-in dashboards and by Oracle BIEE (or any other BI platform) for customized reporting.

Blueprint Concepts:

Logical modelling drives the entire MDM implementation. The model contains the following:

- Structure of the entities and attributes, as well as the relationships between these.
- Rules for data quality, standardization, enrichment, match and merge.
- Design or hierarchies, forms, tables and complex objects for accessing and manipulating the data
- Policies for data security and retention.
- Workflow definition for data stewardship.

From these definitions in the model, all the technical aspects of the hub are generated:

- Internal or public (staging and publishing) table structures as well as web services from the structure.
- Certification process generated in SQL from the declarative quality, enrichment, etc. rules.
- User interfaces from the hierarchies, forms, tables, etc., secured by the policies.
- Activities managed in the user interfaces from the workflow definition.

Blueprint Timeline:

1. The designer creates the logical model, using a web frontend. The model is stored in the Metadata Repository.
2. The designer deploys the model in a MDM hub, generating the technical artifacts.
3. ODI or SOA publish data from source system into the MDM Hub using generated staging tables or web services. The hub converts this source data into golden data through the certification process.
4. ODI or SOA consume golden data using the web services and public tables generated from the model, and distribute golden data to the target applications.
5. Users access the data and perform stewardship or authoring operation in the MDM hub through the applications and workflows generated from the model.
6. Convergence Pulse loads automatically the Pulse Metrics data, which are used by Oracle BI EE for advanced reporting.

Conclusion

Master Data Management is a complex challenge that requires a solution made for the business and with the business. As a value driven initiative, it requires an approach and a platform that uses the existing components of the enterprise technical ecosystem while keeping a strict focus on the data governance aspects of the initiative.

Contact address:

FX Nicolas

Semarchy

14A, Rue Professeur Deperet

69460, Tassin-La-Demi-Lune

Phone: +33(0)4.78.83.05.30
Email fx.nicolas@semarchy.com
Internet: www.semarchy.com