A Metamodel for Master Data

Baba Piprani[1] and Suneil Dham[2]

[1] MetaGlobal Systems, Canada
[2] Office of the Superintendent of Financial Institutions, Canada
babap@metaglobalsystems.com, SDHAM@osfi-bsif.gc.ca

Abstract. The term 'Master Data' brings up different interpretations and connotations, especially with vendors espousing the "single version of the truth". Is there a single version of the truth? Practical realities suggest that we need to continue to live with existing versions – at least until the "single version" is reached. So how do we handle the co-existence of "multiple versions" of the truth? This paper examines a metamodel that defines what master data is, the types of master data, and criteria for determining master data. The metamodel visits master data administration, privileges and master data services including change management, metadata migration and data migration for master data is also reviewed. The impact of changes to an enterprise going the master data route, along with the emphasis to influence future management direction is also examined. Several administration aspects make use of established ISO standards in metadata IS:11179-3 and IS:19763.

Keywords: Master Data, metamodel, ORM, metadata, ISO 19763, ISO 11179, Metadata Registry, MDM.

1 Why Metadata First?

Data Processing, Web Services and Electronic Data Interchange rely heavily on accurate, reliable, controllable and verifiable data recorded in databases or in some persistent state. A prerequisite for correct and proper use and interpretation of data is that both users and owners of data have a common understanding of the meaning and representation of the data as per the Helsinki Principle [1]. To facilitate this common understanding, a number of characteristics, or attributes, of the data have to be agreed upon and defined. These characteristics of data are known as "metadata", that is, "data that describes data". The ISO/IEC 11179-3 Metadata Registry standard [2] provides for the attributes of data elements and associated metadata to be specified and registered as metadata items in a Metadata Registry.

Metadata, or data about data, provides administrators and business users with descriptions of the data or informational objects that they can access.

There are two types of metadata; technical metadata and business metadata. Both types of metadata are important in application software deployment, construction, maintenance, and the use of a data warehouse or data mart.

Technical metadata is used by administrators and software tools, and provides the technical descriptions of data and operations. Technical metadata includes information about source data, target data, and the rules that are used to extract, filter, enhance,

R. Meersman et al. (Eds.): OTM 2010 Workshops, LNCS 6428, pp. 447–456, 2010.

cleanse, and transform source data to target data. Technical metadata could be created by a relational database management system, by data warehouse and transformation tools, or by the data warehouse operations personnel. Examples of these include database statistics, descriptions of transformations, scheduling, etc.

Business metadata is used by business analysts and end users, and provides a business description of informational objects. It assists end users in locating, understanding, and accessing information in the data mart, data warehouse, or other informational sources. Business metadata might include the calculation used to create a particular value, the data and time a report was created, or a description of the approval status of the projected forecast.

In other words, metadata is a general term for data that describes information. The information so described may be information represented in a computer system; e.g. in the form of files, databases, running program instances and so on. Alternatively the information may be embodied in some system, with the metadata being a description of some aspect of the system such as a part of its design. This metadata may describe any aspect of a system and the information it contains, and may describe it to any detail and rigour depending on the metadata requirements of the organization.

2 So Let's Get to Data, and Master Data…

All businesses use data. While we discussed metadata in the paragraphs above, the business is more concerned with the 'truth' value of the data it operates with. Agreed, we need both metadata and data as a type-instance pair to operate in a sane and normal environment. This is where several problems arise. Data is assumed to be sacrosanct since our friendly and reliable computer system is giving it to us, right? Not so! The truth of the matter is that in more than 50% of the cases, data quality and data integrity of the business data is questionable, see Gartner [4].

Data duplication and data integrity are major issues confronting IT applications today. It is not uncommon to come across multiple sets of redundant data on customers and other items of primary focus in an organization. A good average figure that is the norm is that there is about 30 to 60 percent redundant data or duplicated data on clients or customers in a typical organization. The search goes on…for the 'single version' of the truth.

Enter Master Data Management…which basically refers to data relevant to the conduct of the business on which transactions or analyses are based. Creating and maintaining good quality master data has become a necessity, and actually is critical today---considering the recent emphasis on regulatory compliance, Service Oriented Architecture (SOA) thrusts, enterprise integration, mergers and acquisitions. In a nutshell, master data is non-transactional data about people, things, places, and concepts [3], typically held in what used to be 'master file' data in legacy systems, circa 1970's. This basically means master data is the organization's non-transactional data, and the supporting reference data. In other words there are basically two kinds of master data – the common reference data like types and categories pertaining to the properties or characteristics of data (e.g. Country code, Address Type), and the other being 'master file' type data like, customers, clients, vendors, products, etc…essentially data that the organization uses for tracking through transactions.

What this means is that every application or service in an organization is concerned with the sanctity of Master Data and transactional data. So let us ask ourselves…what then is the scope of Master Data? Is it just separating out the Customer records, or

Product records etc…namely, business common data on which transactions are based, and the reference tables, or, is it yet another little item to attend to and move along, or, is there something more to it?

This paper establishes a model for Master data, defining its scope via a Process Model Decomposition, then defines ORM [6] schemas as applicable for managing Master Data. This will help analysts and business managers in establishing project resources or scoping the efforts involved in managing Master Data in an enterprise.

3 The Master Data Business Process Model Decomposition

Figure 1 defines a Master Data process model decomposition that highlights typical business processes involved with going the Master Data route. This model is borrowed from current user experience involved in establishing and stabilizing master data hub and associated master data hub services to address Service Oriented Architecture (SOA) requirements.

We feel it is important that this paper discusses in detail the Business Processes P1, P2 and P3 in Figure 1, for which ORM/NIAM [7] metamodels are provided below. The processes P4, P5 and P6 will need to be addressed during Master Data deployment.

Note that the notation used in the ORM/NIAM schema for Fact Types is using Fact Type Identifiers and omits role names for the purpose of making the diagrams less crowded due to space limitations in this publication. Thanks to Ooriane Corp. for the use of Ooriane Designer and Ooriane Semantic Analyzer for ORM/NIAM graphics.

3.1 Manage Master Data Definition (P1.1) - Define Master Data

It is best at this stage to get our mindset to think both "type" and "instance" for Master Data as a 'married pair'. In other words, when we talk about Master Data, it means both the metadata and the data associated with it.

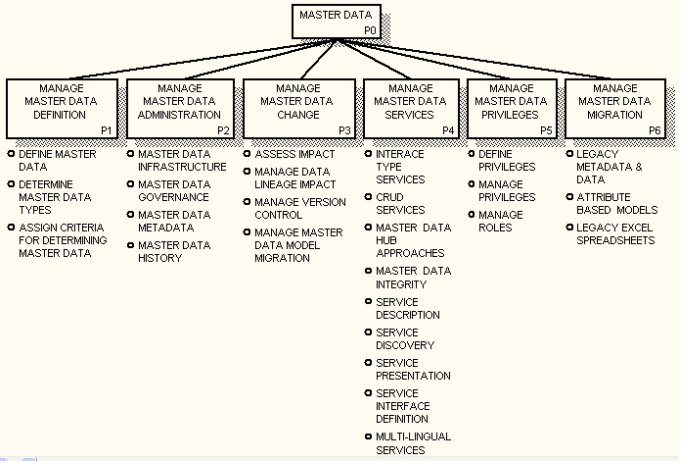

Fig. 1. Master Data Process Model Decomposition (IDEF0 notation, BPWin)

Firstly, we need to define data relevant to the conduct of the business on which transactions or analyses are based, e.g. the basic business tombstone data.

Then we need to identify common data and data roles that are shared across business functions, e.g. Country code used at head office and also used at branch location.

Next is common metadata shared across business functions. E.g. common semantics like customer definition, client definition. It is here there is a major ROI return and payoff in establishing a metadata registry concerning the essential semantics of the data relevant to the business model of the enterprise. A data definition template was designed for use to collect metadata about master data artifacts, called 'Form 19'. An example metamodel of an extract of a "Form 19" used at the Office of the Superintendent of Financial Institutions (OSFI) in Canada is depicted in Figure 2. This Form 19 is based on the ISO 11179-3 Metadata Registry Standard.

3.2 Manage Master Data Definition (P1.2) - Types of Master Data

As mentioned previously, Master Data pertains to Reference Data e.g. country code, address type etc., and, Non-Transactional Data, e.g. employee position, contact information, client information etc.

3.3 Manage Master Data Definition (P1.3) - Criteria for Determining Master Data

Some of the essential requirements or criteria for determining master data are:

- Data or data roles that are shared over multiple functional areas
- Need a single consistent version with standard terminology, definitions and business rules (NOTE: there may be multiple versions to begin with…plan for migration to a single version and strategy to co-exist and transition in the interim)
- Willingness to collaborate on centralized governance of the selected data.

3.4 Manage Master Data Administration (P2)

It is important to establish accountability and responsibility for managing master data administration---much like there is a role in the organization for data base administrators and data administrators. Items involved in this category are:

- Infrastructure administration
- Stewardship and Trustee
  - Roles and Responsibilities
  - Governance Model
  - Master Data Administration (Models and Data)
- Metadata Management (OSFI Form 19, a subset of ISO 11179-3 Metadata Registry Standard and ISO 19763 Metamodel for Interoperability Standard)

See Figure 2 for a model for metadata definition, as a subset of ISO11179-3.

Figure 2 depicts a Metamodel subset for Metadata Definition for Data Elements in a Metadata Registry for Master Data capture.

The following ORM/NIAM schemas depict the constituent formal schemas for each of the sub-registries referenced in Figure 2.

Figure 3 depicts an ORM/NIAM schema of the metadata artifacts being collected pertaining to business oriented definitions and semantic relationships that would

facilitate drawing up a domain ontology for a Data Element (DE). The relevant definitions are provided in Table 1.

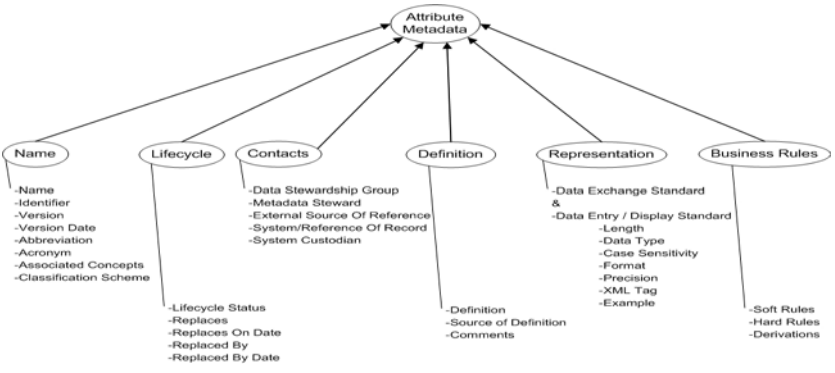

Fig. 2. Metamodel (subset) for Master Data Metadata Definition (Form 19)

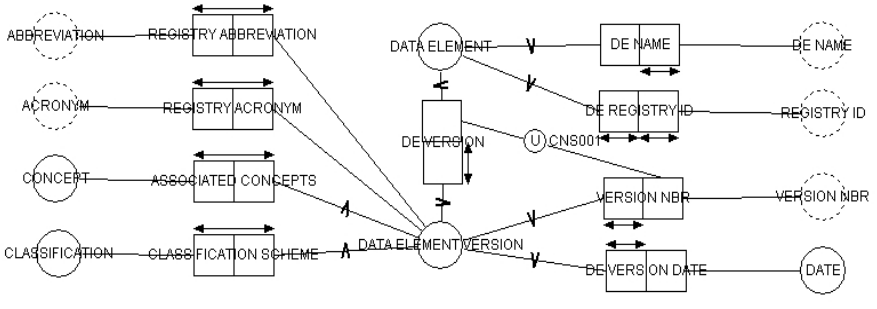

Fig. 3. Metadata Management - Name

Table 1. ORM/NIAM Schema Definitions for Metadata Management - Name

#	Object	Definition
1.1	Name	Business term name in its unabbreviated form.
1.2	Identifier	Any user defined data attribute identifier used for reference purposes
1.3	Version	Version number of metadata definition.
1.4	Version Date	Version date of the version number
1.5	Abbreviation	Recommended short form of the attribute name using standard abbreviations, if any.
1.6	Acronym	Recommended acronyms of the attribute name
1.7	Associated Concepts	Perspective, abstract idea or another attribute with which this data attribute is associated with, or is a property of, or can be mapped to.
1.8	Classification Scheme	Descriptive information for a generic level arrangement of objects into groups based on characteristics that the objects have in common.

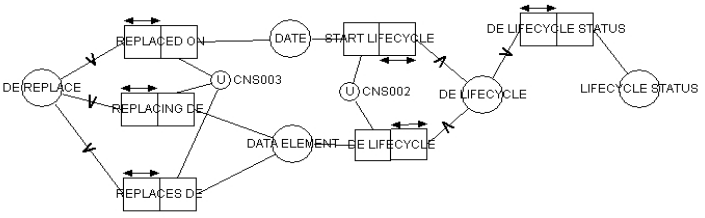

Fig. 4. Metadata Management – Lifecycle

Figure 4 depicts an ORM/NIAM schema of the metadata artifacts being collected pertaining to relevant lifecycle semantics that would facilitate on-going maintenance of metadata for a Data Element. The relevant definitions are provided in Table 2.

Table 2. ORM/NIAM Schema Definitions for Metadata Management - Lifecycle

#	Object	Definition
2.1	Life Cycle Status	Condition of the metadata in its lifecycle (draft, proposed, approved, retired).
2.2	Replaces	Name of attribute (old attribute), which is being replaced by this attribute.
2.3	Replaces on date	Date of replacement for the old attribute
2.4	Replaced by	Name of attribute (new attribute), which replaces this attribute.
2.5	Replaced by on date	Date on which this attribute was replaced.

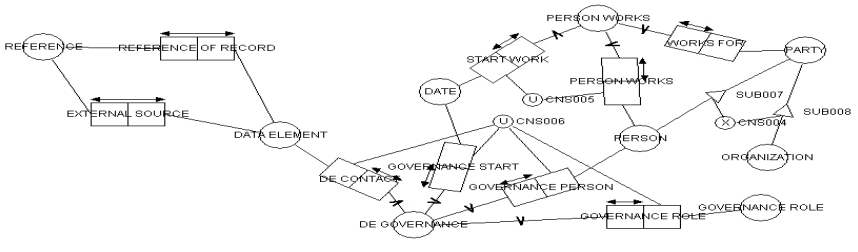

Fig. 5. Metadata Management – Contacts

Table 3. ORM/NIAM Schema Definitions for Metadata Management - Contacts

#	Object	Definition
3.1	Data Stewardship Group	The organization(s) responsible for the accuracy of the attribute's definition.
3.2	Data Steward	The person(s) responsible for the attribute meta-data definition (name, contacts, definition, business rules).
3.3	Source of Reference	Reference number and/or title of an adopted data/metadata standard from an external source.
3.4	System or Reference of Record	The manual or automated system that serves as the authoritative source for accurate data values.
3.5	Custodian	The person(s) responsible for the maintenance and quality of the actual data in the system of record.

Figure 5 depicts an ORM/NIAM schema of the metadata artifacts being collected pertaining to Contact semantics that would facilitate on-going Data Governance requirements of metadata for a Data Element. The relevant definitions are provided in Table 3.

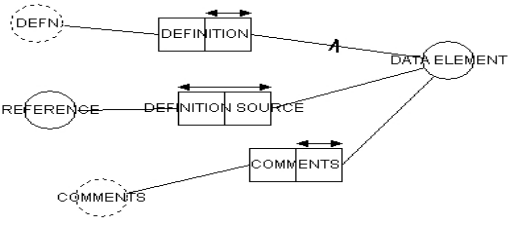

Fig. 6. Metadata Management – Definition

Figure 6 depicts an ORM/NIAM schema of the metadata artifacts being collected pertaining to Definition semantics as per ISO/IEC 11179-4 that would facilitate on-going Data Governance requirements of metadata for a Data Element. The relevant definitions are provided in Table 4.

Table 4. ORM/NIAM Schema Definitions for Metadata Management - Definition

#	Object	Definition
4.1	Definition	The textual description of the attribute as per ISO/IEC 11179-4, see [5]
4.2	Source of Definition	The publication, directive, standard, system, organization or person(s) responsible for developing the attribute definition.
4.3	Comments	Additional information to aid users in understanding the purpose and use of attribute.

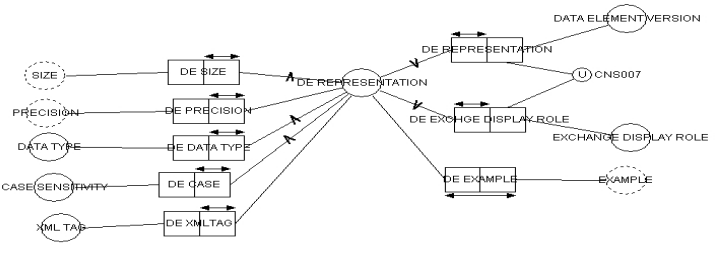

Fig. 7. Metadata Management – Representation

Figure 7 depicts an ORM/NIAM schema of the metadata artifacts being collected pertaining to Data Exchange semantics concerning representation and display of metadata for a Data Element. The relevant definitions are provided in Table 5.

Table 5. ORM/NIAM Schema Definitions for Metadata Management – Representation

#	Object	Definition
5.1	Exchange Display Role	The data element representation mode whether for data exchange or data entry/display formatting purposes
5.2	Size	The maximum & minimum allowable lengths for the raw data.
5.3	Data Type	The kind of data. Examples are: alphabetic, binary, numeric, alpha-numeric
5.4	Case Sensitivity	A specification of whether or not the data is to be upper, lower, or mixed case.
5.5	Format	A specification of the way the raw data should be arranged.
5.6	Scale and Precision (if numeric)	The total number of digits and positioning of the decimal point if applicable.
5.7	XML Tags	Specific XML (Extensible Markup Language) tags associated with attribute
5.7	Example	Provide example(s) of valid raw data and descriptor, if appropriate.

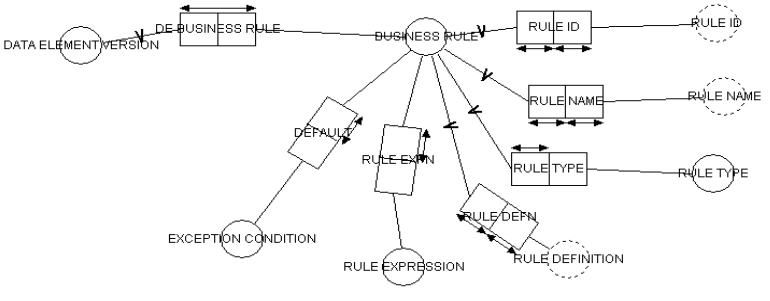

Fig. 8. Metadata Management – Business Rules

Figure 8 depicts an ORM/NIAM schema of the metadata artifacts being collected pertaining to Business Rule semantics for a Data Element. The relevant definitions are provided in Table 6.

3.5 Change Management (P3)

It is important to recognize that changes do occur to both metadata and data pertaining to master data in particular. We need to be able to conduct:

- Impact assessment (e.g. if the data model for master data is changed, what systems are affected)
- Lineage management (e.g. what systems (applications, databases, reports) are using the data)
- Version Control (e.g. versioning of the master data data model)
- Migration (e.g. migration of a new master data model version through deployment environments like Development/Test/User Acceptance/Production as well as data)

Table 6. ORM/NIAM Schema Definitions for Metadata Management – Business Rules

#	Object	Definition
6.1	Business Rules	Contains rules that are business subject area specific and are used to identify the business rule data violators to provide data quality and integrity. Integrity violations on the data will be reported on using this category classification. The corresponding error reports will be forwarded to the data stewards. The Business Rules for definition call for: - Cardinalities of relationships (one-to-zero/one/many, many-many, recursive, parent-child update/delete rules) - Value domains - Super/sub-type - Nullability - Derivations (computed values) - Unique constraints for business keys (not surrogate keys) - Other business rules The Business Rules for Execution call for: - Handling exception conditions - Notification - CRUD sequencing (e.g. can it recognize cardinality rules e.g. insert supertype first then subtype)

3.6 Master Data Services (P4)

The involvement and depth of adoption of Master Data Services depends on the degree of commitment of the organization towards SOA and transitions to the proposed infrastructure and technology. Important in this area are:

- Interface types (SQL, Web Services…)
- Create/Retrieve/Update/Delete/Search
- Federated or Consolidated master data hub approaches
- Managing Referential Integrity between master data and any referencing data
- Service description and discovery
- Service presentation
- Services interface definition
- International Language support

3.7 Privileges (P5)

Security and access are particularly important for when changes or inserts are taking place. However, depending on the nature of business involved, security for read access may become an important issue. Privileges may need to be involved with both at the model level and the data level. Attention needs to be given of privileges over the time continuum based on affects of temporality. Some of the basic functions involved are:

- Define Privileges
- Manage Privileges
- Manage Roles

3.8 Migration of Metadata and Data (P6)

Migration towards Master Data is often overlooked and can involved serious issues when it comes to how the metadata is going to be migrated. Not every vendor offers a smooth and complete metadata transfer. Also needing to be addressed is metadata quality and data quality issues. Some of the issues to be noted are migration of metadata and data from:

- Existing schemas and data to new Master Data registry
- Existing data models
- Existing business rule declarations
- Existing Excel or other spreadsheet ad hocs
- Parallel versions of 'multiple versions of the truth' i.e. the master data version as the single version and, the local legacy version accompanied with mappings during transition, etc.

4 Conclusion

Migrating towards Master Data needs a lot of consideration and planning – in particular: data quality issues, establishing semantic equivalence, co-existence/transitioning. This paper has provided an insight on a model for migrating to a master data environment. Of particular concern, there is widespread confusion and mixed direction in terms of availability of decent migratory software that is of decent industrial strength quality that can be applied with real semantics towards establishing a strong foundation for corporate master data. It is not a simple exercise like purchasing a Master Data package and magic will happen! The encompassing of a SOA based environment will require an engineerable approach to adopting Master Data, including planned phases for the inevitable exercise in data harmonization and data cleansing. Caveat emptor…

References

1. van Griethuysen, J., ed.: Technical Report on Concepts and Terminology for the Conceptual Schema and the Information Base. ISO Technical Report ISO IEC TR9007:1987. International Standards Organization, Geneva (1987)
2. International Standard ISO IEC 11179-3:2003 Metadata Registries, International Standards Organization, Geneva
3. Wolter R., Haselden K.:The What, Why, and How of Master Data Management, http://msdn.microsoft.com/en-us/architecture/bb190163.aspx
4. Gartner Group Report: Gartner Press Release, Gartner Website – Media relations (2005), http://www.gartner.com/press_releases/pr2005.html
5. International Standard ISO IEC 11179-4 Rules and guidelines for the formulation of data definitions, International Standards Organization, Geneva
6. Halpin, T., Morgan, T.: Information Modeling and Relational Databases, 2nd edn. an imprint of Elsevier. Morgan Kaufmann Publishers, San Francisco (2008) ISBN: 978-0-12-373568-3
7. Nijssen, G.M., Halpin, T.A.: Conceptual Schema and Relational Database Design. Prentice Hall, Victoria (1989)