The Influence of Master Data Management on the Enterprise Data Model

For DAMA_NY

Tom Haughey
InfoModel LLC
868 Woodfield Road
Franklin Lakes, NJ 07417
201 755-3350
tom.haughey@InfoModelUSA.com
Agenda

• Definitions
  – Enterprise Data Model
  – Master Data
  – Reference Data
  – Master Data Management
  – The future of MDM
  – MDM and products

• The Context of Enterprise Information Mgt
  – The framework
  – Enterprise Info Management
  – Types of data
    • Metadata
    • Reference
    • Master data
    • Transactional data
    • Historical data

• The EDM
  – Types of
  – Uses of
  – How to build (top down, bottom up)
  – Why EDM is important to MDM

• MDM Usage Process
  – The MDM process
  – General usage types
    • Operational
    • Analytical

• MDM implementation Styles
  – Consolidation
  – Registry
  – Coexistence
  – Transactional Hub
  – Comparison with the ODS

• The ODS
  – Definition
  – Characteristics
  – Comparison of MDM and ODS
  – Passive ODS
  – Active ODS
  – View of data movement

• Conclusion
  – Trends
  – MDM Ecosystem
Objectives and Introduction
The Need for Master Data Management

- Consider a customer who has taken out a mortgage
  - The bank sends mortgage solicitations to that customer
  - Why? Because the customer information used by marketing is not consistent with the customer information in customer services
- A common reason is growth through mergers or acquisitions.
  - Two organizations which merge will typically create an entity with duplicate master data (since probably each has its own data)
  - Theoretically, DBAs resolve such duplication as part of the merger
  - In practice, because applications can have dependencies of their own data, a special reconciliation process is used instead
- Over time, more mergers and the problem multiplies.
  - Data-reconciliation processes become unmanageable
  - Firms have many separate, poorly-integrated master databases
  - This affects customer satisfaction, operational efficiency, decision making and regulatory compliance
- This problem is further accentuated by complex heterogeneous IT environments, shrinking budgets, and fast changing business conditions.
Definition of Master Data

- **Master Data** is the consistent and uniform set of objects, identifiers, attributes and rules that describe the core (and relatively stable) entities of the enterprise and that is used across multiple business processes.
Master Data Management (MDM) Defined

• A business capability that enables an organization to:
  – Identify trusted master data
    • Defines and/or derives the most trusted and unique “version” of core enterprise data (e.g., customer, product, employee, asset, material, location, etc.)
    • Usually, this enterprise data is captured, maintained, and used across disparate systems and business processes for operational and analytical uses
  – Leverage master data to improve business processes and decisions
    • Incorporate this master version of the data within functional business processes (sales, marketing, finance, support, etc.) that will provide direct benefit to employees, customers, partners, or other relevant stakeholders within an organization.
  – How the data will be consumed by other applications or systems within the context of a business process provides the most value

Source: Forrester
Dimensions of MDM

Methods of Use
How do we consume master data:
Collaborative, Operational,
Analytical, etc.

Architectural Style
How do we design the solution
(coexistence, transaction,
registry, etc.)

Entities
What master data do we need
(Customer, Product, etc.)

Implementation
How to start, how much to
do at once

MDM Process
Master Data
Reference Data

• Reference data is commonly defined as “any kind of data that is used solely to categorize other data found in a database, or solely for relating data in a database to information beyond the boundaries of the enterprise”, according to Malcolm Chisholm

• Typically includes data that is commonly called codes, status and role entities

• There are four kinds of reference data, namely:
  – Things external to the enterprise, such as country codes.
  – Type codes, status codes, and role codes, such as Customer Type, Order Type.
  – Classification schemes, such as market segment classifications and industry classifications.
  – Constant values, often externally supplied, such as tax rates.
Other Annoying Acronyms

• EAI – Enterprise Application Integration
  – The use of software and computer systems design principles to integrate a set of enterprise computer applications.

• EII – Enterprise Information Integration
  – The process of providing a single interface for viewing all the data within an organization, and a single set of structures and naming conventions to represent this data.
  – The goal of EII is to get a large set of heterogeneous data sources to appear to a user as a single, homogeneous data source.

• CDI – Customer Data Integration
  – The combination of the technology, processes and services needed to create and maintain an accurate, timely, complete and comprehensive representation of a customer across multiple channels, business lines, and enterprises.
  – Typically there are multiple sources of associated data in multiple application systems and databases.

• PIM – Product Information Management
  – Processes and technologies focused on centrally managing information about products, with a focus on the data required to market and sell the products through one or more channels.
MDM and Technology

- MDM is not just about technology
- Vendors will try to sell you this but first and foremost MDM is about common meaning and sharing stable data
  - MDM technology can help organizations achieve and maintain a single view of master data across an entire enterprise, enabling business and IT initiatives to perform in better unison, allowing for opportunities to increase revenue, reduce costs, achieve effective compliance, reduce risk and improve business agility
  - The technology is still fairly young
  - It continues to mature and has great potential for enabling more-efficient and effective business processes in the future
  - In addition, vendors will sometimes try to sell CDI or PIM as MDM
MDM Is More Than CDI or PIM

- MDM had common roots in two different parts of the business: how to achieve single view of product and customer
  - “Single view” doesn’t mean one place to look
  - We mean, “from a single source of master data, each user in context to the task at hand, can perceive the information in the way they need too” which really results in multiple views from one source
- How vendors talk about MDM’s relevance
  - Too many relate MDM to CDI only, as if that were the only source.
  - Some, but fewer, relate MDM to PIM only
  - The reality is both (and even more) are parts of MDM’s history. As PIM and CDI, these two technologies did start at different times, and in different industries, but they are both part of MDM
- MDM includes the whole space of customer and product, and more
- Besides, technology is not the point
- MDM is not just about data quality, not just about cost, not just about timeliness. At the end of the day, this is about business efficacy
The Context
Levels of Data

- Three levels of decisions

Realm of Analytical Modeling

Strategic

Tactical

Operational

Master/Reference

Unstructured

Structured

History
Context of Master Data Management

• All the extended components of a data model are needed for full MDM
Facets of Data

- **Source**
  - Raw data elements from original systems
  - Selection of candidates for shared data
  - Example: 7 systems use Account Status

- **Semantic**
  - Meaning in terms of definition, stewardship, processes and rules
  - Multiple Account Status created for different systems

- **Structure**
  - Business relationship rules and granularity
  - Interrelationships and mapping of data
  - Account Status is different in values, meaning, data type in each, such as hybrid data types (account status plus sales channel)

- **Domain**
  - Set of values and characteristics at logical level, and actual storage and values at physical level, which can vary vastly by system
  - MDM needs to provide a standard set of Account Statuses that everyone can use plus a standard set of rollups
Types of Data

- **Metadata**
  - A Customer is ..... 

- **Reference**
  - Customer Status defines ...

- **Master data**
  - Customer, Product, ...

- **Transactional data**
  - Order, Delivery, ...
  - Changes are often kept as new line items

- **Historical data**
  - Customer changes over time
  - (Slowly or rapidly) changing dimensions
Components of a Complete MD Data Model

- MDM requires a data model that is complete in all components
- A complete data model consists of these components:
  - Sources
    - **Sources** represents the origins of the data element and its characteristics in the original system
    - Sources are critical to understand for BI applications and DW
  - The model diagram
    - The visual with its entities, attributes and relationships
    - This represents **structure** or syntax
  - Metadata
    - The definitions of everything in the diagram
    - This constitutes its **semantics**
  - Domains
    - Attribute definitions require definition of the **domain**
    - This includes data type, length and set of values
  - Supplementary business rules
    - Any business rule that cannot be structurally represented must be documented in a supplement to the data model
    - This is part of the **semantics**
The Enterprise Data Model (EDM)
**Definition of the Enterprise Data Model**

- An integrated view of the data produced and consumed across an entire organization
  - Represents a single integrated definition of data, independent of any system or application.
  - Also independent of physical implementations, such as how the data is physically sourced, stored, processed or accessed.
  - Represents the objects important to an organization and the rules governing them.
  - It should incorporate an appropriate industry perspective
- Integrated data provides a "single version of the truth" for the benefit of all (but be careful of this expression!!!)
- An EDM is a data architectural framework used for integration
  - It enables the identification of shareable and/or redundant data across functional and organizational boundaries.
  - It minimizes data redundancy, consistency and errors
  - It is essential to data quality, consistency, and accuracy.
- EDM includes **more than** master data
Levels of the EDM

• An EDM can be built at three levels of abstraction
  – The Enterprise Subject Area Model is created first,
  – This can be expanded by subject area to an Enterprise Conceptual Model
  – This can be further expanded, creating Enterprise Logical Models
• Although the models are interrelated, they each have their own unique identity and purpose
• Creating an EDM is as much an art as a science
• An EDM is created in its entirety, relative to the best knowledge available at the time
  – It is an iterative process
  – However, there will always be more revealed
General Use of the EDM

- As a data architectural framework, an EDM is the "starting point" for all data system designs
- The model is used in the same fashion an architectural blueprint is for constructing a building
  - Providing a means of visualization
  - Serves as a framework supporting planning, building and implementation of data systems
  - For enterprise data-intensive initiatives, such as an Operational Data Store (ODS) or Data Warehouse (DW), an EDM is essential, since data integration is the underlying principle for them
- An EDM facilitates integration of data and diminishes the data silos, inherent in legacy systems
Other Uses of the EDM

- It also plays a vital role in several other enterprise type initiatives:
  - Data Quality initiatives
  - Data Stewardship
  - Data System Scalability
  - Industry Data Integration
  - Integration of Packaged Applications
  - Information Strategic Planning
Sample EDM (Subject Area Level)

- Sales example
Sample Conceptual Data Model

- Establish a starting point for detailed data modeling
- Define scope and boundaries of the data model
- Serve as a management communication vehicle
- As a data architecture planning tool
- As part of an enterprise data model
### Characteristics of Models by Level

<table>
<thead>
<tr>
<th>Conceptual</th>
<th>Logical</th>
<th>Physical</th>
</tr>
</thead>
<tbody>
<tr>
<td>• High level</td>
<td>• Detailed</td>
<td>• Optimized</td>
</tr>
<tr>
<td>• General purpose</td>
<td>• Project specific</td>
<td>• Platform specific</td>
</tr>
<tr>
<td>• Major entities only</td>
<td>• All entity types</td>
<td>• Includes auditing</td>
</tr>
<tr>
<td>• No reference entities</td>
<td>• Include appropriate reference entities</td>
<td>• Chooses integrity enforcement</td>
</tr>
<tr>
<td>• Might use keys</td>
<td>• Keys express full identity</td>
<td>• Efficient keys</td>
</tr>
<tr>
<td>• Only major attributes</td>
<td>• Uses natural keys</td>
<td>• Resolves subtypes</td>
</tr>
<tr>
<td>• Many-to-many’s permissible</td>
<td>• May use subtyping</td>
<td>• Some denormalization</td>
</tr>
<tr>
<td>• May use some subtyping</td>
<td>• All attributes</td>
<td>• Some redundancy of data or relationships</td>
</tr>
<tr>
<td>• Two broad versions:</td>
<td>• No many-to-many’s</td>
<td>• Some derived data</td>
</tr>
<tr>
<td>• Project-specific high level data model</td>
<td>• Non-redundant and thereby normalized</td>
<td>• Utilizes DBMS features</td>
</tr>
<tr>
<td>• Cross-project planning model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Method of EDM Creation

**TOP-DOWN MODEL EXPANSION**

- Business Strategies
- Subject Area Model
- General and Industry Knowledge
- Conceptual Data Model

**BOTTOM-UP DATA SYNTHESIS**

- Logical Data Model 1
- Logical Data Model 2
- Logical Data Model 3
- Detailed Individual Knowledge
Two Types of EDM

• Though not generally defined this way, it is possible to create two different EDMs

  – An Operational EDM

    • Which addresses the operational-tactical requirements of the business
    • This is the typical EDM

  – An Analytical EDM

    • Which address the reporting and analysis needs of the business
    • The fact and dimensions, above all the reporting hierarchies, will significantly differ in this model
The Need for an EDM in MDM

• Full MDM is **not possible** without an enterprise view of data
  – You must have a common (and detailed) understanding
  – Only then can applications understand how to share
• So how to express this?
  – As an ontology?
  – As an ER model?
• Our belief is that full MDM is **not practicable** without an EDM, i.e., expressed as an ER or class model
• What is an ontology in philosophy?
  – The study of the nature of being, existence or reality in general, as well as of the basic categories of being and their relations
  – Deals with questions concerning what entities exist or can be said to exist, and how such entities can be grouped, related within a hierarchy, and subdivided according to similarities and differences
• In computer science, an ontology is a formal representation of a set of entities within a domain and the relationships between them
Our View on the EDM for MDM

• Our view is that an ontology is not enough to enable MDM
  – Full practicability requires understanding the identity and grain of each data element
  – This focus is provided by an ER (Entity Relationship) model
• Why?
  – An ontology defines objects and the hierarchy of their relationships
  – Yet, a complete sharing of data requires the understanding of the structure, identity and grain
• Consider the example of Liquid Net Worth
  – What does it mean?
  – “The total part of an party's net worth that can be readily turned into cash. Liquid net worth includes investments such as stocks and mutual funds, but does not include assets that are difficult to readily convert, such as real estate or cars.”
  – Is “the party” a customer, prospect, household, account or position?
  – Does it vary over time?
Grain of the EDM

- The EDM needs to represent a common and application-neutral view of the data
- It should be as detailed as the MDM process requires
- Should the EM have more granularity and refinement of business rules that a business currently requires?
- It is guided by three questions (originally proposed by Peter Drucker) that define its intrinsic qualities
  - What business are we in?
  - What business will we be in?
  - What business should we be in?
- The EDM should be created to address one of these perspectives
- Each of these will reveal significantly different models
- We have seen EDM’s that are more detailed than any business area uses
  - Are such EDM’s overnormalized?
The MDM Process
Types of Usage of MDM

- **Cooperative**
  - Supports the processes for cooperative authoring of master data, including defining, approving, creating and enhancing master data
  - Usually consists of a workflow with manual and automated tasks

- **Operational**
  - An MDM server acts like an OLTP system responding to requests from different users and systems
  - Focuses on providing stateless services in a high performance environment

- **Analytical**
  - Includes a closer tie-in between MD and analytic applications such as BI
  - Can take three forms:
    - Supplying a trusted source of dimensions
    - BI apps performing analytics on MD
    - Analytics being performed by the MD system itself
• There are four general architectural styles for MDM, namely,
  – Consolidation
  – Registry
  – Coexistence
  – Transactional Hub
System of Record / System of Reference

- System of record
  - The primary trusted copy of Master Data
  - The best source of truth
- System of reference
  - A replica of the master data synchronized with system of record
Coexistence Style

- MD is stored in various locations
- Includes a physical golden record instance that is synchronized with source systems
Transactional Hub Style

- Data becomes a system of record
- Updates happen directly to this system
- As updates happen, MD is cleansed, matched and enhanced
- After updates are accepted, they are distributed to using systems
Registry Style

- Provides read-only source of MD for downstream systems
- MD system holds minimal identifier and
- Provides cross-reference to data managed in other systems
Registry Example

Registry Federation

Name, SSNO, Primary Address, NameID1, Privacy Preferences

System A

System B

Name, SSNO, Primary Address, NameID2, Account

MDM Registry

Name, SSNO, Primary Address, NameID1 System 1, NameID2 System 2
Consolidation Style

- Data from various sources is merged into a single managed DB
- En route, the data is cleansed, matched and integrated to provide a single golden record
- Exemplified by the Operational Data Store (ODS)
Operational Data Store (ODS)

- A tactical environment which stores detailed, near-real time results of committed transactions for a certain period of time for immediate reporting needs and which can sometimes be updated by users.
- OLTP system usually store only short-term histories of data, if history at all.
- An ODS usually stores a partial history
  - Maybe just an accounting period
  - Maximum of one year
- If reporting with SOA requires more history than is in OLTP, and ODS may be necessary.
- The ODS needs to solve DQ problems when importing the data.
- While the ODS can be used to stage data for the DW, that is not its primary purpose.
- An “ODS” must have an operational-tactical purpose to be called an ODS and not just a staging area.
Reasons for an ODS

• ODS integrates data, not just consolidates it
  – The resulting model pulls together existing OLTP source models
• An ODS is often created for one or more of three purposes
  – To achieve **consolidated reporting** by integrating data from multiple disparate sources
    • Because modifying the original systems would be too costly
    • Modification of original system would entail significant change to data, application code, presentations and interfaces
  – To achieve **consolidated (even update) processing**
    • Usually because these are existing silo systems
    • Current processing requires disparate access to these silos
  – To do **operational-to-tactical reporting**
    • Because using the OLTP systems is difficult due to disparate data across them
    • Using the OLTP systems is inappropriate because there would be a performance impact on them (and vice versa)
Typical ODS placement

- Sources
- Message Broker
- ETL
- ODS
- DW
- Reporting
- Data Mart
- Updates

MDM Feeds Systems Throughout
Linkages in an Architected Data Environment

- Linkages across OLTP, ODS, DW, DM, BI; Metadata, Information Resource Mgmt, DQ Mgmt, Data Modeling

Continuous Business Improvement

- Legacy Systems
- OLTP Systems
- External Sources
- XML Sources
- Message Queues
- Other Sources

Operational Data Store (ODS) → Data Warehouse (DW) → Data Marts (DM)

ETL

Master and Reference Data Management
Meta Data Management
Information Resource Management
Data Quality Management
Data Modeling
## Active vs. Passive ODS

<table>
<thead>
<tr>
<th>Passive ODS</th>
<th>Active ODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contains duplicate copy of data</td>
<td>Common &amp; pertinent data resides only in ODS</td>
</tr>
<tr>
<td>Populated in batch</td>
<td>OLTP shares this common data with ODS</td>
</tr>
<tr>
<td>Batch synchronization between ODS and OLTP</td>
<td>Nearly synchronous connection between ODS and OLTP</td>
</tr>
<tr>
<td>Update propagation back to OLTP</td>
<td>Real-time</td>
</tr>
<tr>
<td>Near real-time</td>
<td></td>
</tr>
</tbody>
</table>
Active vs. Passive ODS (continued)

**Passive**
- Customer Data
- EAI Messages
- Batch

**Active**
- Customer Data: Golden Copy
- Real-time Access

**SOA: Customer Maintenance Service**

**Customer Data**
Summary and Conclusion
MDM Ecosystem

MDM Core Components
- CDI \ PIM
- Hierarchy Mgt
- EDM
- DQ
- Metadata
- External data enhancements

Upstream
- ERP apps
- Web apps
- Legacy apps
- External data

Change Management
Data Governance  Org Alignment  Business Activity Mgt

Downstream
- DW
- ODS
- Data marts
- BI tools
- Dashboards
- Portals

Data Movement
ETL, Messaging, etc.
Trends in MDM

• **DM Trend No. 1:** Leveraging data quality initiatives
  – Extending data quality tools to an MDM hub environment
  – Identify, match and store data without having to cleanse the source
  – Leverages data cleansing rules already established
  – Maximizes data quality investments

• **MDM Trend No. 2:** Replatforming
  – A common platform for multiple data subject areas
  – For instance, integrating multiple servers onto a single platform
  – This reduces the number of physical platforms while leveraging existing skill sets

• **MDM Trend No. 3:** Building solid business cases for MDM
  – Transcend the "feeds and speeds" conversation
  – Propose bona-fide business value of MDM such as case management in state government
    • Track an individual despite multiple identities and addresses
    • More quickly target food stamp fraud, thereby saving millions

Source: Jill Dyche