TopQuadrant Webcast with Malcolm Chisholm
March 18, 2015

“The Foundations of Successful Reference Data Management”
Today’s Program

Introduction of Agenda and Speakers

I. Foundations of Successful Ref. Data Management
   - What is reference data and why is it important?
   - Challenges of reference data management
   - What are some best practices for governance and management
   - What capabilities should you look for in a reference data solution?

II. Introducing TopQuadrant’s new offering for RDM
   - TopBraid Reference Data Manager
   - Designed to address core capabilities identified by Malcolm to support a modern RDM solution
   - Short walkthrough of tasks that support the life-cycle of a reference dataset

III. Questions and Answers
TopQuadrant Company

Focus:

- TopQuadrant was founded in 2001
- Our focus is to harness emerging technology to build practical but innovative business applications.

Foundation:

- We continue our strong commitment to standards-based approaches to data semantics

Our Mission:

- Empower people —by making enterprise information meaningful
TopBraid Solutions

TopBraid Enterprise Vocabulary Net™ supports collaborative management of enterprise metadata, business glossaries and taxonomies used in search, content navigation and data integration.

TopBraid Insight™ is a semantic virtual data warehouse that enables federated querying of data across diverse data sources as if they were in one place.

TopBraid Reference Data Manager™ supports the governance and provisioning of reference data including the curation of reference datasets (code-lists) with comprehensive metadata.
Reference Data in Context

Increasing:
- Per Value Data Quality Importance
- Semantic Content

Increasing:
- Volume of Data
- Population Later in Time
- Shorter Life Span

Metadata
Reference Data
Transaction Structure Data
Enterprise Structure Data
Transaction Activity Data
Transaction Audit Data

Most Relevant To Design
Most Relevant To Outside World
Most Relevant To Business
Most Relevant To Technology
# Layers of Data

| Metadata          | The data that describes all aspects of an enterprise’s information assets, and enables the enterprise to effectively use and manage these assets.  
|                  | Here it is confined to the structure of databases. Found in a database’s system catalog. Sometimes included in database tables. |
| Reference Data    | 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.  
|                  | Codes and descriptions. Tables containing this data usually have just a few rows and columns. |
| Transaction       | Data that represents the direct participants in a transaction, and which must be present before a transaction fires.  
| Structure Data    | The parties to the transactions of the enterprise. E.g. Customer, Product. |
| Enterprise        | Data that permits business activity to be reported and/or analyzed by business responsibility.  
| Structure Data    | Typically, data that describes the structure of the enterprise. E.g. organizational or financial structure. |
| Transaction       | Data that represents the operations an enterprise carries out  
| Activity Data     | Traditional focus of IT – in many enterprises the only focus. |
| Transaction       | Data that tracks the life cycle of individual transactions.  
| Audit Data        | Includes application logs, database logs, web server logs. |
Importance of Reference Data

- Data Quality problems have widespread impact
- Lack of understanding leads to bad business decisions
- The same table occurs in many different applications
- 20-50% of the tables in a database are Reference Data
- Unnecessary and difficult mappings for data integration
- Need it to understand the world outside the enterprise
- Need it to turn data into business information
Central Reference Data Management Unit

Why Specialized?

<table>
<thead>
<tr>
<th>Reference Data has unique properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g. It has meaning, and is added to production environments</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference Data has unique challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g. It has to be synchronized across many applications</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference Data has unique risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g. It is often misunderstood leading to “miscodings” etc.</td>
</tr>
</tbody>
</table>

Why Centralized?

<table>
<thead>
<tr>
<th>Need for standardization</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g. Which Country Code will be used – GENC, ISO Alpha-2, ISO Alpha-3…</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Need for one place in enterprise to deal with external authorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g. Who ensures we get the NAICS updates</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Need to set up governance for internal reference data mgmt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g. How are Customer Type, Product Line managed?</td>
</tr>
</tbody>
</table>

There are a number of reasons why enterprises should set up a central unit for RDM.
There are many tasks that a central Reference Data Unit (RDU) must perform for External Reference Data.

Some of these tasks could be performed outside the RDU but all must be governed by the RDU.
• Typically Internal Reference Data tables are managed poorly and have no governance
• Governance is needed to assign accountabilities and enforce standard processes that drive up quality
Producers of Internal Reference Data may be well governed, but both Internal and External Reference Data can be misunderstood, misused, and abused in operational environments. This impacts downstream use, data integrity across the enterprise. Governance is required.
• Producers of Internal Reference Data may be well governed, but both Internal and External Reference Data can be misunderstood, misused, and abused in operational environments.
• This impacts downstream use, data integrity across the enterprise.
• Governance is required.

<table>
<thead>
<tr>
<th>Enterprise</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Reference Data Producers</td>
<td>Central Reference Data Unit</td>
</tr>
<tr>
<td>Reference Data Table</td>
<td>Govern</td>
</tr>
<tr>
<td>Govern</td>
<td>Govern</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example: Customer Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why isn't Corporate Customer in this table?</td>
</tr>
<tr>
<td>I think Goldman Sachs is a Retail Bank…</td>
</tr>
<tr>
<td>Hmm…no code for Hedge Fund – I’ll put one in</td>
</tr>
<tr>
<td>Accts Recvbl</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accts Recvbl</th>
<th>Order Entry</th>
<th>Sales Data Warehouse</th>
<th>Treasury</th>
</tr>
</thead>
<tbody>
<tr>
<td>I’ll use the code for Asset Manager to book Private Equity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hmm…no code for Hedge Fund – I’ll put one in</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This Hub is for both the production and distribution of Reference Data.
Data Stewards typically produce some data in the Hub.
The Hub may source some (usually most) Reference Data from legacy applications (typically transaction applications).
• Production of Reference Data is done in specialized environments.
• Only production-ready Reference Data is placed in the Hub.
• All other environments subscribe to Reference Data from the Hub.
Summary and Capabilities to Consider in Solutions

- Profile an External Authority
- Profile an External Reference Dataset
- Support Semantic Analysis of each Element in Reference Dataset
- Document Semantic Analysis
- Import Reference Data into a Repository
- Assign Accountabilities for RDM Tasks
- Track Changes to Reference Data
- Support Distribution of Reference Data

- There are other capabilities to consider in Reference Data solutions, but these are fundamental.
Summary and Capabilities to Consider in Solutions

- Profile an External Authority
- Profile an External Reference Dataset
- Support Semantic Analysis of each Element in Reference Dataset
- Document Semantic Analysis
- Import Reference Data into a Repository
- Assign Accountabilities for RDM Tasks
- Track Changes to Reference Data
- Support Distribution of Reference Data

- There are other capabilities to consider in Reference Data solutions, but these are fundamental.
A flexible web-based solution for governing and provisioning reference data in the enterprise:
– Governance
– Provisioning
– Comprehensive metadata
– Enrichment

TopBraid Reference Data Manager™ (TopBraid RDM) makes it easy to bring consistency and accuracy to reference data management and use.
TopBraid RDM enables more meaningful and effective use of reference data by capturing and managing **semantic metadata** about **reference data** and also about **reference datasets**.
Who Benefits from RDM?

TopBraid RDM has a variety of users including:

• **Data stewards** whose primary responsibility is governance of reference data

• **Subject matter experts** who contribute to identification and development of reference data and advise the data stewards

• **Data managers** who consult RDM to find reference data suitable for their applications

• **Application administrators** and **database administrators** who create and use RDM exports in order to load reference data into the systems they are responsible for maintaining

• **Business analysts** who want to better understand the meaning of reference data as they design business reports

• **End users of reference data** who want to make sure that they are using correct codes and contribute knowledge that may enrich the body of information managed by RDM
Today’s Example

Data Steward: dataset governance

• Creating a Reference Dataset
• Importing Reference Data
• Maintaining Reference Dataset Metadata
• Modifying Reference Data
• Exporting Reference Data
• Extending Business Concept Model (Ontology) to support additional properties

(Role = Manager)
Demo
Collaboration features

• Change management, versioning and working copies
• Task assignment
• Security and permissions
• RACI support: Responsible, Accountable, Consulted, Informed
### RACI Support

<table>
<thead>
<tr>
<th>Role</th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responsible</td>
<td>North America's Data Steward</td>
</tr>
<tr>
<td>Accountable</td>
<td>Chief Data Officer</td>
</tr>
<tr>
<td>Consulted</td>
<td>Information Strategy Program Manager</td>
</tr>
<tr>
<td></td>
<td>VP of Business Analytics</td>
</tr>
<tr>
<td>Informed</td>
<td>Flight Search Application Owner</td>
</tr>
<tr>
<td></td>
<td>Flight Tracker Application Owner</td>
</tr>
</tbody>
</table>
Additional features

• Data quality and validation rules
• Hierarchy management
• Crosswalks
• APIs, integration and customization
• Model driven flexibility for present and future needs
• Empowers data stewardship – easy maintenance – minimal IT involvement
• User friendly web-based UI
• Metadata capabilities
• Easy customization
• Governance

For more information:
rdm-info@topquadrant.com