USING TAXONOMIES AS MASTER REFERENCE METADATA FOR THE ENTERPRISE

April 11, 2012

Jeannine Bartlett – Chief Solutions Architect, Earley & Associates

Data Management Association San Francisco Chapter (SF DAMA)
Presentation

Agenda

1. *A Parallel Universe*: Taxonomy vs. MDM - Enterprise Perspectives

2. *A Brave New World*: Convergence of Structured & Unstructured MDM

1. A PARALLEL UNIVERSE

Taxonomy vs. MDM
Enterprise Perspectives
Parallel Perspectives

• What is the difference between a data architect, a taxonomist and an information architect?

• Data architects are concerned with *structured* data and technical aspects of applications and database design.

• Taxonomists are concerned with *unstructured* content classification, semantics and the meaning of terms.

• Information architects consider how structured data elements, unstructured content meaning and user intent combine to form the user experience.
The (Traditional) Data Architect

View of Metadata

Operational Metadata
- Data source description
- Data fields
- Data structure
- File description
- Database definition (e.g. ddl)
- External files lineage

ETL – Metadata
- Rule descriptions
- Data cleansing
- Data extension
- Data transition

Database Design, Physical Rules
- Modeling data
- Security
- Access / update performance
- Cube definition
- Aggregation

Source data → Metadata → ETL Tools → Target database

Extract/Transform/Load

Modeling Tools

BI Metadata
- Norms for report creation
- Report-Folder
- Report-Format
- Report-Execution
- Data mining process
- Rollup, pivot tables, OLAP

BI/Reporting Tools

Copyright © 2012 Earley & Associates, Inc. All Rights Reserved.
The (Traditional) Taxonomist View of Metadata

- Taxonomy: system for organizing and classifying content
- Metadata: information about our content, housekeeping, as well as semantic and structural information
- Content Objects: groups of metadata that are assembled into components that are then assembled into pages or documents

How will taxonomy surface on the front-facing application?

What do the wireframes suggest?

How do people interact with content?

How does the content architecture deliver the front-end design?
Best Practices for Taxonomy Design

- Clear grouping principles
- Categories sufficiently distinct
- Sub-categories are logical fits for their parents
- Is polyhierarchy used? Is it consistent?
- Is the structure consistent across categories?
- Do top level categories convey business domains and processes?
- Are single-child categories avoided?
- At what levels are product attributes exposed? Global attributes used where appropriate?
Difficulty Exchanging Paradigms

“We have everything we need to know about our products from the SKU…”
_**Merchandiser**_

“Taxonomy values are just a list of terms…”
_**Data Architect**_

```
Caption = BDY FATHER FRM DAUG
```
Occasion  = Birthday  Recipient  = Father  Sender  = Daughter

“We can drive content semantics from composite database keys…”
_**Database Architect**_

“I’m more comfortable indexing content than dealing with back end systems…”
_**Taxonomist**_

<table>
<thead>
<tr>
<th>Occasion</th>
<th>Tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birthday</td>
<td>Funny</td>
</tr>
<tr>
<td>Wedding</td>
<td>Romantic</td>
</tr>
<tr>
<td>Halloween</td>
<td>Heartfelt</td>
</tr>
</tbody>
</table>
2. A BRAVE NEW WORLD

Convergence of Structured & Unstructured MDM
Taxonomy Frameworks in Enterprise Architecture

Case Example: Motorola’s Global Taxonomy Framework serves Multiple Applications

Copyright © 2012 Earley & Associates, Inc. All Rights Reserved.
Taxonomy is a Semantic Classification Structure

- **Diseases**
  - Alopecia
  - Bulimia
  - Cancer
  - Diabetes
  - Epilepsy

**Ontology**

- **Taxonomy**
  - **Hierarchy of Terms**
    - at least one level deep
    - parent-child relationships
    - node reference id’s for access and persistence
  - **Hierarchy of Concepts**
    - synonyms
    - language variants
    - metadata attributes
    - attribute inheritance
  - **Associative Relationships**
    - many-to-many relationships
    - named relationships have semantic meaning
  - **Object Classes**
    - strong object class typing
    - concepts vs. entity instances
    - Semantic Web reasoning

**Controlled Vocabulary**
- a list of words
- no classification structure
- limited use for drop downs, dictionaries

**Complexity & Enrichment**
Taxonomy | At minimum, a hierarchy of terms, but... 

... describe processes, content use cases, expressive nuance, not just topics

- define major categories and preferred terms to *improve precision*
- enrich with synonyms and language variants to *improve recall*
- enhance with best bets and other associative relationships to *improve discovery*
Hierarchical & Associative Relationships

Content Choreography Views - Semaphore Ontology Manager

Hierarchical

<table>
<thead>
<tr>
<th>Type</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT</td>
<td>Assembly Template</td>
</tr>
<tr>
<td>contained in Advertisment</td>
<td></td>
</tr>
<tr>
<td>contained in Blog Post</td>
<td></td>
</tr>
<tr>
<td>contained in Call to Action</td>
<td></td>
</tr>
<tr>
<td>contained in Event Calendar</td>
<td></td>
</tr>
<tr>
<td>contained in Promotion</td>
<td></td>
</tr>
<tr>
<td>contained in Review</td>
<td></td>
</tr>
<tr>
<td>contained in Testimonial Quote</td>
<td></td>
</tr>
<tr>
<td>contained in Title</td>
<td></td>
</tr>
</tbody>
</table>

Associative

<table>
<thead>
<tr>
<th>Type</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>sent to chan... Blog &amp; eNewsletter</td>
<td></td>
</tr>
</tbody>
</table>

Equivalence
Taxonomy as Lifecycle & Rights Metadata
Metadata Attributes in a Taxonomy Framework
Taxonomy to Drive Content Choreography

Site Search Facets
- UX Page Context
- Personalization
- Site Navigation
- Site Search Experience
- Transient SEO & Campaigns
- Social Content

UX, Navigation, Campaigns, SEO

Site Search Ingestion & Indexing
- Templates
- Content Authoring
- Digital Asset Creative
- Translation & Localization
- Content Variants
- Content Lifecycle

Content & Digital Asset Creative, Translation & Assembly

Metadata
- WCMS Search
- WCMS Content
- WCMS Digital Assets
- Taxonomy Mgmt System
- Metadata Mgmt & MDM
- Change Control

Development, Deployment & Operations Readiness

Copyright © 2012 Earley & Associates, Inc. All Rights Reserved.
Taxonomy as Master Reference Metadata

- Business Intelligence needs metadata on structured data as the basis for *quantitative* analysis, taxonomy on unstructured content for the *results of* analysis.

- Extend traditional *quantitative* BI with *qualitative* BI from unstructured content once a taxonomy is applied:
  - analysis of risk/no risk claims by disease category
  - analysis of call center issues by product to monitor a recall
  - analysis of ad placement effectiveness by social media context

- Master Data Management needs taxonomy – both aim to define a “single version of the truth”
  - MDM to eliminate *structured data redundancy*
  - Taxonomy to eliminate *unstructured content ambiguity*
Challenges of Changing Metadata Paradigms
Enterprise Taxonomy Management Tools

Existing taxonomies, thesauri, controlled vocabularies imported and converted

Taxonomy Manager
Subject matter experts from around the organization design and maintain taxonomies

Taxonomy Server
- Metadata assignment in content, digital asset, document and records management systems
- Automated tagging, indexing and search facets
- Structured data, business applications, data warehousing and BI
- Website and portal navigation, Intranet directories
- Content syndication, subscription, alerts, monitoring

API calls, batch export, Taxonomy Connectors
Terminology Challenges

Digital Asset Management
- Structure and terminology to support non-text asset location and reuse

E-commerce suite
- Reconcile vendor product metadata with structure and format for catalog, merchandising, order management

Guided navigation
- Facets and attributes based on taxonomy resolve with search user experience best practices

Content management
- Content and document types, topics/subjects, audiences, etc. to support unstructured information

Product taxonomy

Cross sell/up sell
- Semantic relationships for related products, controlled terminology for merchandisers to support specific promotions
Different Workflows & Clock Speeds

www.bestbuy.com

Brick and Mortar

Product taxonomy

BI/Reporting

Finance

Signage

Ad Campaigns
Common Governance

Questions

1. How do you pick the source of truth?
2. How do you integrate the practice of merging reference data into the enterprise model when adding a new application? When is it done? Who does it?
3. Is it just about having a metadata repository that holds the descriptions? Is that what we mean by integration? Or do we mean actually changing systems to be harmonized?
4. How often do changes in the organization or regulatory requirements mean that harmonization efforts get thrown out? How do you prevent that?
5. Is there such a thing as a dynamic metadata layer?
Taxonomy Change Management Scenarios

**Regular changes:**
Inserting new terms (e.g. New concept)

*Renaming terms (e.g. to better fit user language)*
Splitting terms (e.g. Category too large, split into 2)
Merging terms (e.g. Too few items in sub-categories)
Retiring a term (e.g. Term obsolete)

**Major changes:**
Adding a new facet (e.g. New offering type)
Retiring a facet (e.g. Cessation of activity)

IMPORTANT: what are the technical and organizational effects of change?
When Taxonomy Terms Change, What Happens to Metadata?

- Reference ID Persists When A Value Changes
  Manage term changes by reference id and value

OPTION 1. By Reference and Value
- Date created: May-15-2009
- Document name: IRESSA Recommended...
- Asset Type ID: 111 (reference)
- Asset Type Value: Brochure

OPTION 2. By Value
- Date created: April-15-2010
- Document name: IRESSA Recommended...
- Asset Value: Pamphlet
- Asset Type: IRESSA

Renaming Term “Brochure” to “Pamphlet”
# Benefits Of Persistent Reference ID’s

<table>
<thead>
<tr>
<th>Term Identification and mapping of related terms to Reference ID persist</th>
<th>Follows database design standard to insure referential integrity</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Preferred Term CAN be reflected in multiple systems consistently via Reference ID</td>
<td>A user could change the name, but integrated systems will not be compromised because the GUID is still the same</td>
</tr>
<tr>
<td>Avoid retagging and re-indexing</td>
<td>No implications against legal hold documents</td>
</tr>
</tbody>
</table>
Governance Conclusions

• Metadata and taxonomy integration cannot be accomplished in a vacuum
• Technology stack and infrastructure will determine how metadata needs to flow
• Development of master taxonomy needs to be integrated with master data management programs
• Governance initiatives (data governance, content governance, taxonomy governance) describe organizational policies for triggering and implementing change
• Downstream impact on systems and processes needs to be carefully evaluated
SUMMARY

- Successful Enterprise MDM initiatives leverage data architect & taxonomist resources, processes & perspectives
We design and deliver content management, search and information supply chain solutions for global organizations and their customers.
Contact Us

Jeannine Bartlett
VP & Chief Solutions Architect,
Earley & Associates

Jeannine.Bartlett@earley.com

www.earley.com