

Taxonomy and Classification for RIM

Bruce Schinkelwitz, MAS CRM
Presented to Greater Los Angeles ARMA, Nov. 18, 2009

Agenda

- Introductions
- The RIM Challenge
- Taxonomy & Classification, a Primer
- RIM Taxonomy
- Developing Effective Classification

The RIM Classification Challenge

- Why is it so hard to develop RIM classifications?
- Why are RIM Classifications not widely used (if at all)?
- File Plan hierarchies are the RIM 'organizing principle'
- We are concerned with control of 'outputs,' rather than shaping the 'architecture' and use of information
- We are handicapped by the traditional construct of retention schedules as file plans

Definitions

- Taxonomy:
 - *“a classification of things or concepts, as well as to the principles underlying such a classification” (Wikipedia)*
 - *“the rules or conventions of order or arrangement” (Lambe)*
- Classification:
 - *A system of organizing things or concepts into groups or categories, or taxonomies, using established criteria*

Taxonomies are Easy

- They are natural to human existence
- They are individual, collective
- They are pragmatic
- We intuitively understand taxonomy as a 'language'
- We use information retrieval taxonomies every day:
e.g., the Web

- Fun: 1 minute exercise

Taxonomies are Very Difficult

- For Users
 - They are individual
 - Shared Drives, CD collection, books, etc.
 - They are situational/contextual
 - By activity, responsibility, viewpoint
- For Creators
 - End users are unhelpful taxonomists - own focus
 - Taxonomies have to address multiple points of view
 - Taxonomies have to accommodate change, the unknown

Taxonomy is

- A form/subset of a Classification Schema
 - Groups related things - many kinds of relationships
 - Schema can be informal and ad hoc (our exercise), or formal (Dewey)
- Semantic
 - Relationship of terms - a vocabulary
 - Concepts must be known and meaningful to users
- A knowledge map
 - Users immediately grasp the structure and anticipate where to find items

Characteristics of Good Taxonomy

- Knowable
 - Users intuitively know/expect a rule governs the taxonomy
- Consistent
- Predictable

Main Taxonomy Types

- Lists
- Trees
- Hierarchies
- Thesauri
- Facets

Taxonomy Type: Hierarchies

- Organized strictly as parent-child, dominant-subordinate
- Characteristics:
 - Inclusiveness - top level includes all subordinates
 - Relational consistency - same relationship distance between levels
 - Inheritance - child entities inherit all characteristics of parent
 - Mutual exclusivity - an entity can belong in only one class
- Eliminates ambiguity
- Very difficult to build and maintain because of rigidity
- Common examples: ERMS systems, File Shares
 - Why ERMS systems are often limited in success as 'enterprise' systems
 - Why File Shares are a mess!

Types of taxonomies: Thesauri

- Like a dictionary, but organized by subject or concept with scope notes
- Organized
 - BT/NT; Related terms; Use/Use for terms
- Example: [AGIFT Thesaurus](#)
- Seattle City Clerk Thesaurus:
 - <http://clerk.ci.seattle.wa.us/~public/thesaurus/newtoc.htm>

Types of taxonomies: 'F-type'

- We know that things can have more than one commonality, or one arrangement
 - *A book can be about >1 subject*
 - *Items in a store can be arranged in many ways*
- "F" taxonomies arrange things by commonality other than parent-child
- Example: [Wine AVA Classification](#)
- Example: [Amazon](#)
- F= 'facet': THINK: METADATA

Functional Classification, or Business Classification Schema (BCS)

- Represents business activity as FUNCTION - ACTIVITY, linked to retention
- Focuses classification on the *context* of the records
- Benefits for RIM:
 - Stability
 - Connects activity to information flows
 - Provides context beyond the single transaction
 - Allows for application of retention at point of records creation, not 'retirement'
 - Reduces the effort in the development, revision and management over time of retention schedules

Developing Effective Functional Classification - 1

- Classification schemas are almost always a mix of taxonomy types
 - Lists, Trees, Hierarchies, Facets
- General principles for Hierarchies:^{*}
 - A List is unusable above 10-15 items - break it down
 - As the classification grows deeper, the purely hierarchical structure becomes very hard to use
- Faceted taxonomy ('metadata view') allows much more content coverage
 - Facet = 1 'view' of the content (e.g., all records by a date)
 - BUT: really only workable with ERMS or Taxonomy system

^{*}Source: Lambe, 190-191

Developing Effective Classification - 2

- Focus on Business activities/processes: [Sample HR Map](#)
- Gather evidence of what users are doing:
 - Vocabularies
 - Categories
 - Organizing principles
- Document 'Sources' for each activity
 - what is the legal or policy basis?
- What are the records series (Retention schedule rules) for each document in the activity?
- Who are the process stakeholders?

Developing Effective Classification - 3

- Develop a matrix for core activity & records metadata
 - For each record 'need,' what is the term/value that binds?
 - E.g., for finding Personnel files, what is the value you would search by?
- Develop a Functional Thesaurus
 - A 'dictionary' of business activity terms in your taxonomy
 - Guide for Users
 - Filing
 - Searching
 - Supplement with terms (subjects) not in the classification but which users use - point them to the 'Use' term
- Aim for this: [Sample RIM Classification](#)

Developing Effective Classification - 4

- How to deal with the Retention Schedule as File Plan challenge?
- Typical ERMS systems permit only retention applied to a folder, with one disposal rule per folder
 - Personnel files have different retention rules by document type, forcing the creation of separate folders per doc type:
 - HR
 - Personnel
 - Bruce S. file
 - Bruce S. hiring letter - CY + 7
 - Bruce S. disciplinary action - CY + 3
 - Bruce S. Training - CY + 2
 - Bruce S. Separation - CY + 75

Reference: RIM and Standards

- ISO 23081 - Metadata
- Dublin Core - ISO 15836
- ISO 15489 - Records management
- ARMA/ANSI
- DIRKS
- These are important because
 - They structure information
 - They structure processes related to the management of information
 - They define approaches that rely on taxonomy and classification

Reference Library

- Patrick Lambe. *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness*. Chandos, 2007.
- State of New South Wales. *Strategies for Documenting Government Business: The DIRKS Manual*. 2007. <http://www.records.nsw.gov.au/recordkeeping/dirks>
- ISO. Information and documentation - The Dublin Core metadata element set, 15836:2003
- ISO. Information and documentation -- Records management processes -- Metadata for records, 23081:2006.
- ISO. Information and documentation -- Records management, 15489:2001
- ARMA: Controlled Language in Records and Information Management, 2008.
- <http://www.slais.ubc.ca/RESOURCES/indexing/database1.htm>
- National Archives of Australia. Overview of Classification Tools for Records Management. 2003

Thank You.

Bruce Schinkelwitz, MAS CRM

bruceschinkelwitz@me.com