



SLA Taxonomy Division

Selecting software to manage taxonomies and ontologies

Ralph Tamlyn
ralph.taxonomy@gmail.com
rtamlyn@us.ibm.com

June 2013

Copyright © Ralph Tamlyn 2013

1



Who is Ralph Tamlyn

- Lead taxonomist for IBM's CIO Enterprise Solutions and Web Enablement team
- Responsible for unifying IBM taxonomies for all internal and external web content.
- Over 10 years working with classification metadata, taxonomies, and content integration
- Over 35 years with IBM
- Ralph is in the IBM CIO organization, not sales, not marketing





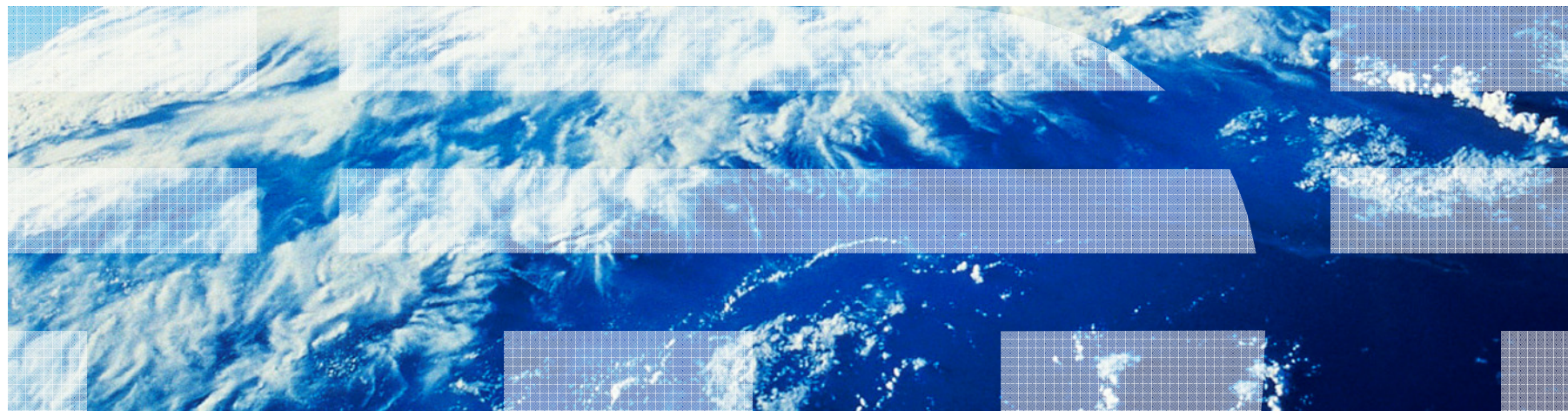
Ralph Tamlyn

rtamlyn@us.ibm.com

ralph.tamlyn@gmail.com

April 2013 SLA Taxonomy Division Webinar

Best Practices for Classifying Enterprise Information



What is Ralph working on?

- Multi-year project started late 2010 / early 2011 to improve the integration and delivery of IBM web content through better management of taxonomies and better management of classification metadata
- The Taxonomy Project was expanded in late 2012 to address taxonomies (controlled vocabularies) used for all information, including information in transaction systems.



Taxonomy Project Objectives

- Enhance the user experience when taxonomies are being used (and adversely affecting users)
- Reduce instances of overlapping taxonomies
 - Identify overlaps
 - Reconcile differences
 - Distribute common taxonomies distributed use
- Facilitate content integration



Sources of Taxonomy needs

- Needs when using taxonomies
- Needs when publishing taxonomies to be used
- Needs when managing taxonomies
- Using, publishing, managing ontologies
- “Non-functional” needs



Taxonomy Topics Omitted

- This presentation does not discuss several important and often complex topics:
 - taxonomy quality
 - no overlapping categories, no classification ambiguity
 - no multiple fact categories
 - child categories are types of parents, not just related
 - reconciliation of overlapping taxonomies
 - process integration
 - governance



Using a taxonomy

- Basic information
 - Category Titles and Codes
- Beyond Basic capabilities
 - Automated consumption (into systems)
 - Subsets for different uses and systems
- Advanced capabilities
 - Delta files
 - User Facing Taxonomies (for Ontologies)
 - Sequential classification (for Ontologies)



Using a taxonomy - Details

- Basic information
 - Category Titles and Codes
 - Best Practice: use codes to allow title changes without needing to touch data; codes also help with category title translation
- Beyond Basic capabilities
 - Automated consumption
 - Subsets for different uses and systems
 - Acceptable for new content – multiple authoring systems
 - Acceptable for current content – may include values no longer used for new content
 - Defining values allowed to pass through a filter
 - Hierarchies – only when appropriate, not just convenient
 - Sequence control
 - Version identification



Using a taxonomy - Details

- Advanced capabilities (also leading to Ontologies)
 - Delta files
 - Help systems assess change impact in advance
 - Help systems react to changes after they occur
 - Automated consumption of delta files
 - User Facing Taxonomies – break the rules for information management but meet the needs of users
 - Sequential classification – the choice of value ABC in a starting taxonomy means a choice should be made from subsequent taxonomy



Publishing a taxonomy

- Simple answer: publish everything necessary to enable the intended uses
- Basic publication
 - Category Titles and Codes
 - Manage translated category titles
- Beyond basic publication needs – properties!!!
 - Filters based on properties to create subsets
 - Include properties in published files when needed in downstream systems
 - Sequence control
 - Versions



Formats in Publishing a Taxonomy

- Human friendly reports and formats
 - Spreadsheets and / or HTML reports
 - Validation of taxonomy and properties
 - Browsing
- Feeds for automated consumption
 - XML, JSON, or CSV
- Subsets: properties to include or exclude nodes
- Include selected properties for the nodes



Advanced Publishing Capabilities

- Delta files
- Real-time visualization capabilities
 - Visualization of changes in context within the taxonomy
- Reports for purposes beyond routine consumption
 - List defective categories / nodes with proposed remediation
 - Show taxonomy changes including actions, e.g deactivation actions



Managing a Taxonomy

- Basic properties - Category Title and Code
- Properties to match usage and publication needs
 - Properties at the Node level / Category level
 - Properties at the Taxonomy / Vocabulary level
- Additional Properties for Taxonomy Management – e.g. owners, business rules



Managing a Taxonomy – Property Classes

- Boolean: true / false
- Free form text
- Numeric
- Value from an allowed values list
- Pointers and hyperlinks to unlimited additional information
 - Other tools
 - Information repositories
- Dates and TimeStamps



Managing a Taxonomy – Property Details

- Boolean true false
 - Subsets
 - Whether or not a defect assessment has been performed
 - Defect presence
 - Designated as strategic taxonomy (this is taxonomy level)
- Numeric
 - Sequence control – sequence may be different for more than one intended use
- Free form text (and strings)
 - Version identifier
 - Defect remediation explanation
 - Complex deactivation guidance
 - Text descriptions, explanations, examples of use



Managing a Taxonomy – Property Details

- Value from another taxonomy
 - Status for new content: active, inactive, proposed
 - Status for old content: active, inactive
 - Deactivation action (simple actions) – when a node or category is deleted or deactivated
 - Ownership – organization or business function
 - Node to taxonomy properties enable Sequential Classification
 - <<Many situation specific examples>>
- Links and pointers to other tools and repositories
 - Pointer to entry in IBM Business Glossary, where the concept is defined – taxonomy level property
- Effective dates, obsolescence dates



Manage a Multiple Facet Taxonomy

- Define a collection of single facets or concepts taxonomies as elements in a larger multiple facet taxonomy
 - Created in response to the needs of a consumer
 - Publish separate files or as a single compound file.
- Note: personally I do not advocate compound files (because isolated changes require complete republication), but this is a decision which should be left to the application owners.



Taxonomy Management Activities

- Workflow within the tool
 - Define states
 - Draft
 - Approved (and therefore frozen)
 - Archived or retired
 - Actions to move between states
 - Approve
 - Archive
 - Create new version – creates a draft
- Convenient work folder – tasks
- Change request management



Taxonomy Change Notifications

- Basic
 - New version has been published – the most basic notification
 - Changes will be coming (“next” draft created)
- Advanced notifications
 - Generally: notification of specific changes in advance: to generate an impact assessment
 - Adds – usually no impact
 - Splits – variable impact
 - Deletes / deactivations – include deactivation action
 - Title modification – “should” have no impact
 - Property modification – simple example: addition to a subset



Advanced Management Function

- Apply business rules to taxonomies
 - Prefixes and codes
 - Simply insert a standard prefix
 - Check for standard prefix (check for typos)
 - Automatically generate next code with designated prefix



Manage a Taxonomy – User Interface

- Direct management of a taxonomy
 - Node / Category level
 - Display and edit all details (properties) for a node or category
 - Display and edit nodes and categories in a tabular format – enhances visualization and editing of properties (not likely useful for text properties)
 - Filters to control the nodes displayed
 - Display and edit taxonomy level details and properties
- Quickly locate a taxonomy
- Quickly locate a node or category in a taxonomy



Manage a Taxonomy – System Interfaces

- Interfaces enabling integration of taxonomies into systems and applications
- Import and APIs to load and modify taxonomies
 - Node addition / insertion / modification
 - Property updates
- Export and APIs to publish taxonomies
 - Filters to provide only the information needed
 - Transforms into formats needed



Function to Manage an Ontology

- Mappings / Relationships
 - Node in taxonomy A to node in taxonomy B
 - Node in taxonomy A to nodes in taxonomies B and C
 - These mappings enable User Facing Taxonomies
 - Node in taxonomy A to the entirety of taxonomy B
 - These mappings enable Sequential Classification Ontologies
 - Taxonomy A to taxonomy B (?)
 - Imports, Exports, and APIs
- Properties on Mappings and Relationships
 - Most frequently used: value from an allowed values list



Advanced Ontology Management Function

- Apply business rules to Ontologies
 - Mapping or Relationship
 - If a node in a taxonomy becomes inactive, the next version of the mapping should neither map to nor from that inactive node.
 - Properties on mappings or relationships can trigger many output variations



Function to Publish an Ontology

- Understand consumption needs
- Publish individual taxonomies
- Publish mappings, often with the taxonomies
- Define and publish integrated collections of taxonomies and mappings to create the ontology
 - Simplest example is two taxonomies and the mapping between them; simplest HTML might portray the relationship visually as parent / child
 - Ontologies are likely to include many taxonomies and mappings



Non-functional Needs

- Backup
- Software environment
 - What other middleware is required?
 - What operating environment is required?
 - What is the application integration environment?
- Access to information in a separate runtime repository versus real-time access to taxonomy management system



Help Consumers of a Taxonomy

- Authentication and Access Control
 - Integration with standard security software (LDAP standard)
 - Control
 - Modifications of a taxonomy (or mapping)
 - Creation of a taxonomy (or mapping); or a new version
 - Deletion of a version or entire taxonomy (or mapping)
 - Approval of individual changes
 - Approval of new versions
- Publication schedule / frequency – business decision
 - Periodic is friendlier to consumers because it is predictable
 - On demand creates complications but is certainly more responsive
 - Fundamentally, applications are responsible for coordination



Examples of IBM's Consuming Systems

- Web Content Management systems
 - Lots of teams developed lots of such systems
 - Teams are being guided to the common solution, but they only switch when they can afford to do so
 - Internal for every business function
 - External for every business unit
- Transaction systems
 - Customer Information
 - Opportunity Management
 - Fulfillment
 - Data warehouses for business and financial reporting



IBM Solution

- IBM evaluated offerings from several vendors, including a services development project
- Solution selected:
 - IBM Reference Data Management (RDM, started as a service offering, now a component in IBM's Master Data Management (MDM) suite)
 - Builds on IBM InfoSphere (UDB/DB2) and WebSphere
 - Integration with
 - IBM Business Glossary
 - ILOG for business rules
 - BPM for advanced workflow
 - Extensions to meet more needs (may get into products)



How does the solution match needs?

- ✓ Usage and Publishing
 - ✓ Basic
 - ✓ Beyond Basic
- ❖ Advanced Usage & Publishing
 - ❖ Reports and Delta Files
 - ❖ Multiple Concept Taxonomies
- ✓ Managing Taxonomy
 - ✓ Basic and Advanced Property Classes and Details
 - ✓ Node level
 - ✓ Taxonomy level
 - ❖ Manage & Apply Business Rules
- ❖ Workflow
- ❖ Notifications
- ❖ Interfaces
 - Business User Visualization
 - ❖ User and System Interfaces
- ✓ Manage Ontology
 - ✓ Mapping level properties
 - ❖ Manage & Apply Business Rules
- ❖ Publish Ontology
- ❖ Non-functional needs

✓ Green - complete
❖ Green – on the way

❖ Yellow – satisfied with design
 Red – design not yet complete

