NASA Taxonomy: A Way to Unify the NASA Information Space

ISKO Singapore
KOS create order & makes sense of things

Ursus Wehrli. The art of clean up: Life made neat and tidy. (http://www.fubiz.net/2011/08/31/the-art-of-clean-up/)
<table>
<thead>
<tr>
<th>Purpose</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>Translate user queries into information retrieval indexing vocabulary.</td>
</tr>
<tr>
<td>Consistency</td>
<td>Enable complete and consistent attribute values.</td>
</tr>
<tr>
<td>Semantics</td>
<td>Specify semantic relationships between and among terms.</td>
</tr>
<tr>
<td>Browsing</td>
<td>Enable users to navigate hierarchies and browse categories to locate content items.</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Aid to help users think about how to search for content.</td>
</tr>
</tbody>
</table>

## Principles of vocabulary control

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eliminate ambiguity</td>
<td>Ensure that each term has only one meaning</td>
<td>Drum (container) vs. Drum (musical instrument)</td>
</tr>
<tr>
<td>Control synonyms</td>
<td>Identify preferred label for each context. Concept vs. label</td>
<td>IBM vs. International Business Machines</td>
</tr>
<tr>
<td>Establish relationships among terms</td>
<td>Equivalence, hierarchy and associative relationships</td>
<td></td>
</tr>
<tr>
<td>Test, validate and maintain terms</td>
<td>Query logs and content analytics</td>
<td></td>
</tr>
</tbody>
</table>
Using warrant to select terms

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literary warrant</td>
<td>The label that most commonly appears in publications (based on natural language).</td>
</tr>
<tr>
<td>Organisational warrant</td>
<td>The official label (based on organisational needs, priorities or policies).</td>
</tr>
<tr>
<td>User warrant</td>
<td>The label users most commonly use.</td>
</tr>
</tbody>
</table>
KOS Schemes: Simple to Complex

Controlled Vocabulary

- A system for identifying and naming things, and arranging them into a classification according to a set of rules.

Synonym Ring

- An arrangement of knowledge usually enumerated, that does not follow taxonomy rules. E.g., Dewey Decimal Classification.

Simple Taxonomy

- A list of preferred and variant terms.

Classification Scheme

- A formal naming and definition of the types, properties, and interrelationships of the entities that exist for a particular domain.

Thesaurus

- A set of words/phrases that can be used interchangeably for searching

Faceted Taxonomy

- A tool that controls synonyms and identifies the semantic relationships among terms.

Ontology

- Taxonomic metadata, or a set of attributes with distinct controlled vocabularies, and semantic relationships among terms and attributes.

Semiotic Schemes

- A formal naming and definition of the types, properties, and interrelationships of the entities that exist for a particular domain.

A list of preferred and variant terms.

After: Amy Warner. *Metadata and Taxonomies for a More Flexible Information Architecture*
What is a taxonomy?

A taxonomy is a particular form of controlled vocabulary in which the labels are organised according to a hierarchy.
Origins of faceted classification

- Mathematician/librarian S.R. Ranganathan (1920s)
- Developed as an alternative to Dewey Decimal System for books.
- "Colon Classification" facets
  1) Personality – topic or orientation
  2) Matter – things or materials
  3) Energy – actions
  4) Space – places or locations
  5) Time – times or time periods

S.R. Ranganathan. Painting by A. Ramakrishna, Art teacher, K.V. No. 2, Vijayawada
(http://www.thehindu.com/multimedia/dynamic/01548/12isbs-ranga_G4_12_1548490e.jpg)
What are taxonomy facets?

- Discrete branches of a taxonomy.
- Consistent, extensible sets of attributes for labeling content and content components.
- Data values for structured data records (or metadata) that allows unstructured content collections to be processed like a database.
- Taxonomic metadata.

Facets = Metadata (with Controlled Values)
Facetted classification

- Categorises items into multiple taxonomies or “stackonomies” based on unique but pervasive characteristics such as geography, type, price, etc.
  - Wines by region \textit{France} > \textit{Alsace}
  - Wines by type \textit{White} > \textit{Chardonnay}
  - Wines by price

- Reflects the domain of the items and allows assignment of \textit{multiple classifications} to items.
  - By region, by type, by price, etc.

- Intended for searching with multiple terms in combination (post-coordination), one from each facet.

[Image of a wine bottle with a map and text: Different types of wine published by Wine Folly.](http://visual.ly/different-types-wine)
How is a taxonomy different from a library cataloging?

- Taxonomy is a post-coordinated indexing system
- It relies on named entities (people, organisations, locations, etc.)

Pre-coordinated system: Each book has its place on the shelf

Post-coordinated system: Items can be in many categories
Case study: NASA Taxonomy goals and objectives

- Make it easy for various audiences to find relevant information from NASA programs quickly
  - Provide easy access for NASA Web resources.
  - Information integration for unified queries and management reporting.
  - Improve search results targeted to user interests.
  - Enable the ability to move content through the enterprise to where it is needed most.
  - Facilitate Records Management and Retention Requirements.
- Be ready to participate in federal XML projects.

Taxonomy goal = Enable knowledge discovery
NASA Taxonomy project timeline

Work Package 1
- Audit Content practices
- Identify & survey stakeholders

Work Package 2
- Build community of practice
- Agree on comprehensive branches & taxonomy detail

Work Package 3
- Test & validate Taxonomy

Work Package 4
- Dublin Core mapping
- Metadata specification
- XML schema development

Follow-on Work
- Integrate with applications
- Long term maintenance

2002

Phase 1

2003

Phase 2

2004+

Follow-on Work
Phase I activities (2002): Fact finding & community building

- Review existing taxonomy terms in light of:
  - Mission and Organizational changes
  - Use cases
  - New content
- Identify potential Gold Sources and subject matter experts
- Make revisions as necessary
- Meet with Review Team regularly for NASA input
Phase 2 activities (2003): Validation & standardisation

- Application of NASA Taxonomy to repositories
  - 5 content sources – LLIS, NODIS, NEPP, OSP, and POLARIS.
  - Map existing metadata schema from test repositories to NASA Taxonomy.
  - Review outcome of faceted navigation test in web environment.
  - Change Taxonomy as needed from test results.

- Develop technical specifications
  - Dublin Core mapping.
  - Metadata specification.
  - XML schema development.
Taxonomy follow-on work (2004+)

- Taxonomy stewardship
  - Maintenance, education and training
  - Facilitate standard adoption process

- Apply in public and internal portals
  - Public and Engineering portals
  - Search integration
    - Verity K2
    - Faceted search and navigation - Seamark
  - Content integration networks for real time delivery

- NASA Enterprise Architecture Group standards development
  - Agency Registries, RSS Syndication, Semantic Web components
### NASA Taxonomy facets

<table>
<thead>
<tr>
<th>Facets</th>
<th>Strategic Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>Overall access filtering</td>
</tr>
<tr>
<td>Access Controls</td>
<td>Sensitivity and access control</td>
</tr>
<tr>
<td>Audiences</td>
<td>Who is the content intended for</td>
</tr>
<tr>
<td>Business Purpose</td>
<td>Why the content was created</td>
</tr>
<tr>
<td>Competencies</td>
<td>What field or discipline is relevant</td>
</tr>
<tr>
<td>Content Types</td>
<td>The genre of the content</td>
</tr>
<tr>
<td>Industries</td>
<td>External partners &amp; businesses</td>
</tr>
<tr>
<td>Instruments</td>
<td>Flight payloads that yield science</td>
</tr>
<tr>
<td>Locations</td>
<td>Sites where work occurs – on and off Earth</td>
</tr>
<tr>
<td>Missions/Projects</td>
<td>NASA’s lines of business</td>
</tr>
<tr>
<td>Organizations</td>
<td>NASA organizations</td>
</tr>
<tr>
<td>Subject Categories</td>
<td>The topic of the content</td>
</tr>
</tbody>
</table>
NASA Taxonomy top 2 levels

**Access**
- Internal
- Public
- US Citizens

**Access Ctrls**
- Classified +
- Public
- Sensitive but Unclassified +
- Unreviewed

**Audiences**
- Educational +
- Employees
- External +
- Government +

**Locations**
- Earth +
- Galaxies +
- Solar System +
- Stars

**Business Purpose**
- Administration +
- Financial
- Management +
- HR +
- Inspector General +
- Legal & Technical +
- Procurement +
- Program
- Formulation +
- Program
- Management +
- Property & Supply +
- Transportation +

**Competencies**
- Bus Knowledge +
- Engineering & Tech Knowledge +
- Leadership & Mgmt Knowledge +
- Mission Operations Knowledge +
- Science Knowledge +

**Content Types**
- Articles, Notes & Papers
- Reviews, Results & Analyses +
- Announcements +
- Calendars & Schedules +
- Case Studies
- Catalogs & Databases
- Correspondence +
- Databases +
- Designs & Specs +
- Drawings
- Educational Materials +
- Forms & Templates +

**Instruments**
- Accelerometers
- Acoustic Sensors
- Anemometers
- Antennae
- Barometers
- Beta Detectors
- Cameras +
- Charged Particle Analyzers
- Cosmic Dust Analyzers
- Dosimeters
- Dust Impact Detectors
- Electrostatic Analyzers
- Faraday Cups +

**Missions & Projects**
- Air Transportation & Aeronautics +
- Aeronautics +
- Astronautics +
- Chemistry & Materials +
- Biological & Physical Research +
- Crewed Missions +
- Data +
- Deep Space & Science Missions & Projects +
- Earth Missions & Projects +
- Ground Facilities +
- Lunar Missions & Projects +
- Solar System Missions & Projects +

**NASA Subjects**
- Aeronautics +
- Astronautics +
- Chemistry & Materials +
- Engineering +
- General
- Geosciences +
- Life Sciences +
- Math & Computer Sciences +
- Physics +
- Social & Information Sciences +
- Space Sciences +

**Organization**
- Commercial Tech Centers +
- NASA Affiliated +
- NASA Centers +
- NASA Contractors
- NASA HQ +
- Other NASA Partners +
- Related Govt Agencies +

**WBS**
- Aircraft / Spacecraft +
- Ed & Public Outreach +
- Ground Systems +
- Launch Vehicle / Services +
- Mission Operations System +
- Payload
- Project Mgmt +
- Safety & Mission Assurance +
- Science / Technology +
- Systems Engineering +
- Systems Integration & Testing +
NASA Taxonomy statistics

- Total number of terms: 5533
  - Preferred terms: 1961
  - Non-preferred terms: 3572
- Hierarchical relationships: 2013
  - Top terms: 12
  - Equivalence relationships: 3718
  - Associative relationships: 0
NASA Taxonomy website


Background & training materials

- Metadata spec
- XML RDF schema
- SKOS files
- Editorial style guide

Taxonomy change process

Links to current vocabularies
Demonstration of guided navigation of test collections

Built using Siderean Seamark (website and software are no longer available)
NASA Taxonomy best practices

❖ Faceted classification schema
  - Modular in nature for easier maintenance.
  - Easy to tag what is appropriate to a use case.
  - Facets give flexibility and power to search.

❖ Polyhierarchy
  - Concepts can appear more than once.
  - Enables knowledge discovery from multiple viewpoints.
  - Fits a user-centric organization.
NASA Taxonomy best practices (2)

❖ Hierarchical granularity
   - NASA taxonomy is broad in nature by design.
   - Different levels of depth depending on attribute set and content.
   - Integration points allow for mapping of local vocabulary terms back to larger semantic framework.

❖ Use of standards
   - Incorporates existing federal and industry terminology standards like NASA AFS, NASA CMS, FEA BRM, NAICS, and IEEE LOM.
   - Complies with metadata standards like Z39.19, ISO 2709 (now ISO 25964), and Dublin Core.
   - Political buy in from CIOs.
Mapping aliases

- Librarian approach allows for a meaningful expression of relationships.
- Synonyms, acronyms, related terms, broader and narrower terms
  - Example: MER A=Spirit, MER B=Opportunity
  - Example: Section 366 = Section 372
  - Example: JIMO = Jupiter Icy Moons Orbiter = Prometheus 1
- Easily maintained in derived RDF files.
- A means of holding institutional memory.
## NASA Specific Metadata Fields

*NASA has data sets unique to its business*

<table>
<thead>
<tr>
<th>Definition</th>
<th>NASA Taxonomy Mapping</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA missions and projects</td>
<td>nasa:missionsProjects</td>
<td>Taxonomy: Missions and Projects</td>
</tr>
<tr>
<td>Business purpose</td>
<td>nasa:businessPurpose</td>
<td>Taxonomy: Business Purpose</td>
</tr>
<tr>
<td>Technical competencies</td>
<td>nasa:competencies</td>
<td>Taxonomy: Competencies</td>
</tr>
<tr>
<td>Standard industry categories</td>
<td>nasa:industries</td>
<td>Taxonomy: Industries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>From NAICS standard</td>
</tr>
</tbody>
</table>
Project outcomes: NASA Taxonomy benefits

… at the NASA Level

- Metadata specification for all NASA content publishers
- RDF schema published with agreed standards (to enable appropriate use and reuse)
- Enhancement of Agency Web publishing processes
- Integration with NASA public portal content management system for:
  - Reduced publishing cycles
  - Better quality of Web materials – coordinated themes
- Integration with NASA Search Engine, Web Site Registration System
- Application in many technical areas, including engineering and science disciplines (STEP and science data dictionaries)
Project outcomes: NASA Taxonomy benefits

... at the Federal Level

- NASA taxonomy development in accordance with e-Gov Act of 2002
- Integration with FEA at the BRM & DRM level
- Increased interoperability with other federal agencies through common data models and standards
- Better interoperability with industry partners for increased speed of mission development
- Enhanced results in First Gov search engine
- Readiness to actively participate in e-Gov initiatives
White Papers and Supporting Documentation

- NASA Taxonomy


Summary

The National Aeronautics and Space Administration (NASA) is composed of more than a dozen separate Centers that engage in a variety of technological activities with missions and projects that reflect their specialized work. Technical communities across the Agency utilize highly evolved engineering and scientific vocabularies that reflect the nature of their disciplines. These vocabularies are often arcane and not easily translated by personnel outside a particular community of practice. This situation has tended to fragment information produced by NASA personnel. The goal behind building and adopting an agency-wide NASA Taxonomy [1] was to develop a consistent framework for handling NASA’s electronic content, and to make it easy for various audiences to find all the relevant information from all the NASA programs quickly.

This talk looks back at a Taxonomy Strategies project to work with NASA from to develop a taxonomy and metadata standard meant to act as a classification scheme encompassing all of NASA Web content, including internal as well as external materials. The taxonomy provided a framework for tagging NASA Web material so that they could be used and reused in different applications such as the NASA portal and its supporting systems, the NASA search engine, and the NASA Web Site Registration System. In 2008, the Taxonomy Strategies was presented with an award from the NASA CIO for “providing major support on the NASA Taxonomy Task with expertise and knowledge on topics of vocabularies and metadata.”

[1] https://www.loc.gov/item/lcwa00099446/