Controlled Vocabularies and NASA Taxonomy Development

Jayne Dutra, Jet Propulsion Laboratory
With Contributions by Joseph Busch, Taxonomy Strategies

SMC-IT
July 16, 2003

Problem Statement

- Workers need unified, universal access to all information, but the real value comes from that portion of the information that actually solves the information problem at hand.

- The amount of time wasted in futile searching for vital information is enormous, leading to staggering costs...

The High Cost of Not Finding Information (2001)

Costs of Not Finding Information

Workers spend up to 2.5 hours a day looking for information

... But find what they are looking for only 40% of the time.

Leveraging Knowledge

Workers spend more time re-creating existing content than creating new content

Communicating

Creating

Searching

Recreating existing content 26%
Creating new content 9%

Image: Workers spend more time re-creating existing content than creating new content.

What Is a Taxonomy?

- Overall scheme for organizing content to solve a business problem
- Representation of a predefined organizational structure that covers a range of subjects in a hierarchical arrangement and that shows correlations between subject areas*
- Optimized site map or information architecture that allows users to intuitively navigate to content

* Taxonomy, Thesaurus, Tagging (2002)

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
- Used to derive metadata that accurately describes target content
What is the NASA Taxonomy?

The classification scheme is meant to encompass all of NASA web content (NASA web space), including internal as well as external material. It is a means for tagging content so it can be used and reused in different contexts.

How to Use the NASA Taxonomy

- This is a generic taxonomy from which specializations can be derived for specific purposes.
- Not all facets need to be used in each instance
- A facet is repeatable
- The taxonomy is modular and dynamic

NASA Taxonomy Basics

NASA Taxonomy Top Level

Information
Audiences
Organizations
Missions and Projects
Industries
Locations
Functions
Disciplines
Chronology

NASA Taxonomy Best Practices

- Design process that:
  - Incorporates existing federal and industry terminology standards like NASA AFS, NAICS, SOC, ACM Classification, and LOM
  - Provides for NASA XML namespace registry
  - Complies with metadata standards like 239.19, ISO 2709, and Dublin Core
- Practices increase interoperability and extensibility

NASA Taxonomy detail

Disciplines
- Science
  - Physics
  - Chemistry
  - Earth Sciences
  - Life Sciences
  - Physical Sciences
  - Engineering
  - Computing and Information Sciences
- Social Sciences
  - Behavioral Sciences
  - Communications
  - Economics
  - Education
  - Human Factors
  - Information Sciences
  - Social Sciences
- Law
- Planning

Disciplines
- Agriculture
- Architecture
- Engineering
Taxonomy and XML

NASA Taxonomy provides controlled vocabularies used to populate elements of more complex metadata schema such as the Dublin Core (www.dublincore.org).

The taxonomy facets map to these schema elements.

---

What are XML schema?

- Data models expressed in XML
- Consistent structure (or syntax) and semantics for XML documents that allow machines to carry out rules made by people
- Names of metadata elements and a consistent set of attribute values or vocabularies for filling them


---

**NASA Taxonomy - Dublin Core Map (Draft)**

<table>
<thead>
<tr>
<th>Dublin Core Element</th>
<th>Definition</th>
<th>NASA Taxonomy Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator</td>
<td>Content maker</td>
<td>Creator,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>creator organization</td>
</tr>
<tr>
<td>Subject</td>
<td>Content topic</td>
<td>subject,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>subject disciplines</td>
</tr>
<tr>
<td>Publisher</td>
<td>Publisher of this</td>
<td>publisher,</td>
</tr>
<tr>
<td></td>
<td>manifestation</td>
<td></td>
</tr>
<tr>
<td>Contributor</td>
<td>Content contributor</td>
<td>contributor,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contributor organization,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contributor organization,</td>
</tr>
<tr>
<td>Type</td>
<td>Genre</td>
<td>type,</td>
</tr>
<tr>
<td>Coverage</td>
<td>Scope, period, date, jurisdiction, etc.</td>
<td>coverage,</td>
</tr>
<tr>
<td>Audience</td>
<td>Content audience</td>
<td>audience,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>STEM audience</td>
</tr>
<tr>
<td>Non DC</td>
<td>NASA missions and projects</td>
<td>NASA missions and projects</td>
</tr>
<tr>
<td>Non DC</td>
<td>Business functions</td>
<td>NASA function</td>
</tr>
<tr>
<td>Non DC</td>
<td>Technical specialties</td>
<td>NASA specialties</td>
</tr>
<tr>
<td>Non DC</td>
<td>Standard industry categories</td>
<td>NASA industries</td>
</tr>
</tbody>
</table>

---

**NASA Taxonomy Benefits**

**... at the NASA Level**

- Metadata specification for NASA content publishers
- Common language for all NASA material
- Integration with OneNASA portal content management system for:
  - Reduced publishing cycles
  - Coordinated message themes by the Agency
  - Better quality of Web materials
- Integration with NASA Search Engine, Web Site Registration System
- Development of XML schema in accordance with DISA Registry (reuse where appropriate)
- Application in many technical areas, including engineering and science disciplines (STEP and science data dictionaries)

---

**NASA Taxonomy Benefits**

**... at the Federal Level**

- NASA taxonomy development in accordance with e-Gov Act of 2002
- Integration with FEA at the 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
The taxonomy addresses the Data Reference Model Layer and enables standardization and mediation.

**Taxonomy Next Steps**

**Test and Validation**
- Confirm stakeholders and communities
- Confirm use case scenarios
- Stratify Test Pilot - Ames

**Dublin Core Mapping**
- Complete Dublin Core mapping
- Create necessary NASA specific tags

**Schema Development**
- Develop XML schema from metadata
- Register schemas in DISA Registry

**Integration with NASA Web Applications**
- Integration with NASA portal and content management system
- Integration with NASA Web Site Registration System
- Integration with NASA web search engine

**Taxonomy Development With NASA**


- Case Study
- Methodology
- Best Practices
- Examples