Enabling Knowledge Discovery
Case Study - Taxonomy Development for NASA

Jayne Dutra, NASA Jet Propulsion Laboratory
With contributions by Joseph Busch, Taxonomy Strategies
October 15, 2003
Problem Statement

- Business 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 by business workers in futile searching for vital information is enormous, leading to staggering costs...

  *The High Cost of Not Finding Information (2001), IDC finding in a study for industry*
Workers typically 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

Searching

Recreating existing content 26%
Creating new content 9%
Finding the right information at the right time to solve the problem at hand
Objectives of Phase One

Interview NASA Knowledge Domain Experts to:

- Understand current strategies and practices for creating, collecting, and organizing information across NASA
- Observe how information is used and organized, the audiences for this information, and the information needs of these audiences
- Elicit goals, hopes, and concerns for an information architecture solution
- Start building a community of interest
Key Findings

- Most (70%) NASA content already has some tagging or is categorized.
- Most (70%) owners add tag content with metadata.
- Almost half (45%) use a standard metadata schema, but many different standards are used.
- Most (60%) use a controlled vocabulary, but lots of different controlled vocabularies are used.

➢ Different NASA constituencies care deeply about what schemas are specified, and what vocabularies are used because ...
## Audiences for NASA Content are Diverse

<table>
<thead>
<tr>
<th>Generalist</th>
<th>Expertise</th>
<th>Specialist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government officials and managers</td>
<td>Educators, education product developers, children 8-12 (and their parents), students K-16+</td>
<td>External vendors, researchers, contractors, grantees, scientists, engineers</td>
</tr>
</tbody>
</table>

NASA employees: engineers, mission controllers, project and task managers, program implementers, functional process teams, contract specialists, administrators, educators, etc.
Audience Uses and Goals Vary Widely

**Admin**
- Better understand the program in total, and obtain scheduling information, project status and best practices.
- Access procurement rules and examples, and procurement action synopses.

**Sci Tech**
- Engineering specifications.
- Scholarly research, competitive intelligence, and general aerospace research.

**Outreach**
- Develop educational products, support current products, learn, etc.
- Topic research and fact finding, topic background research, and downloading curriculum support materials.
- In the classroom as stand alone items, hands-on learning opportunities, class projects, to expand on a student's learning potential.
- Find NASA contact information on services, information about student opportunities, information about career opportunities, and latest educational news.
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 Beta Taxonomy Top Level

- Information
- Audiences
- Organizations
- Missions and Projects
- Industries
- Locations
- Functions
- Disciplines
- Chronology

Taxonomy Best Practices

For Extensibility and Interoperability

- Uses standard facets such as:
  - Disciplines (technical specialties)
  - Functions (of business records)
  - Industries (who you do business with)
  - Locations (on Earth, and off Earth)
  - Organizations (you are part of, affiliated with, or do business with)
  - Projects (your business)
  - etc.

- Re-uses existing standards and vocabulary sources such as:
  - ACM for Computer Science specializations,
  - AFS for Functions,
  - LOM for Educational Roles,
  - NAICS for Industries,
  - SOC for Employee Roles,
  - etc.
Taxonomy Best Practices (2)

For Robustness and Depth

- Provide hierarchical granularity whenever possible
- Reuse the same concept multiple times in the scheme (so that the same concept has multiple parents)
- Map abbreviations, and other aliases as alternate terms or synonyms
- Use standard genre and document type categories in the Information facet
XML Applications

- 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
### NASA Taxonomy - Dublin Core Map

<table>
<thead>
<tr>
<th>Dublin Core Elements</th>
<th>Definition</th>
<th>NASA Taxonomy Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator</td>
<td>Content maker.</td>
<td>dc:creator &lt;br&gt; dc:creator.employee &lt;br&gt; dc:creator.organization</td>
</tr>
<tr>
<td>Subject</td>
<td>Content topic.</td>
<td>dc:subject.organization &lt;br&gt; dc:subject.missionsProjects &lt;br&gt; dc:subject.disciplines</td>
</tr>
<tr>
<td>Publisher</td>
<td>Publisher of this manifestation.</td>
<td>dc:publisher.organization</td>
</tr>
<tr>
<td>Contributor</td>
<td>Content contributor.</td>
<td>dc:contributor &lt;br&gt; dc:contributor.employee &lt;br&gt; dc:contributor.organization</td>
</tr>
<tr>
<td>Type</td>
<td>Genre.</td>
<td>dc:type.information</td>
</tr>
<tr>
<td>Coverage</td>
<td>Space, period, date, jurisdiction, etc.</td>
<td>dc:coverage.locations &lt;br&gt; dc:coverage.chronology</td>
</tr>
<tr>
<td>Audience</td>
<td>Content audience.</td>
<td>dcTERM:audience</td>
</tr>
<tr>
<td>Non DC</td>
<td>NASA missions and projects.</td>
<td>nasa:missionsProjects</td>
</tr>
<tr>
<td>Non DC</td>
<td>Business functions.</td>
<td>nasa:functions</td>
</tr>
<tr>
<td>Non DC</td>
<td>Technical specialties.</td>
<td>nasa:disciplines</td>
</tr>
<tr>
<td>Non DC</td>
<td>Standard industry categories.</td>
<td>naics:industries</td>
</tr>
</tbody>
</table>
E-Government Applications

The Federal Enterprise Architecture is a business-focused framework for cross-agency, Government-wide improvement.

The Federal Enterprise Architecture (FEA) is providing OMB and Federal agencies with a new way of describing, analyzing, and improving the Federal Government and its ability to serve the citizen.

The FEA will eliminate the organizational obstacles that have historically hindered improvement without forcing reorganization.

The FEA is a business-focused approach and is not just for IT.

The FEA provides a common framework for improving a variety of key areas:

Business Line Focus:
- Budget allocation
- Horizontal and vertical information sharing
- Performance measurement and budget/performance integration
- Component Based Architecture

Citizen Centered:
- Cross-agency collaboration
- Improved service to the citizen
- e-Government
- Process integration
- Call center convergence
- and more

JPL Knowledge Management
NASA Taxonomy and the Federal Enterprise Architecture Model

The taxonomy addresses the Data Reference Model Layer and enables standardization and mediation

Federal Enterprise Architecture (FEA)
by permission of Bob Haycock, OMB

Performance Reference Model (PRM)
- Government-wide Performance Measures & Outcomes
- Line of Business-Specific Performance Measures & Outcomes

Business Reference Model (BRM)
- Lines of Business
- Agencies, Customers, Partners

Service Component Reference Model (SRM)
- Service Layers, Service Types
- Components, Access and Delivery Channels

Data Reference Model (DRM)
- Business-focused data standardization
- Cross-Agency Information exchanges

Technical Reference Model (TRM)
- Service Component Interfaces, Interoperability
- Technologies, Recommendations

JPL Knowledge Management
NASA Taxonomy at the NASA Level

Benefits of Taxonomy Development

- Meta data specifications for all NASA Web publishers
- Development of XML schema in accordance with DISA Registry (reuse where appropriate)
- Enhancement of Agency Web publishing processes
- Integration with NASA public 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
Benefits of Taxonomy Development

- NASA in compliance for taxonomy development as specified by e-Gov Act of 2002
- Integration with FEA at the DRM level
- Increased interoperability with other federal agencies through common data models and standards
- Enhanced results in First Gov search engine
- Readiness to actively participate in E-Gov initiatives
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
Blue Sky: Beyond Taxonomies

JPL Knowledge Base

- Standardized project content and data architecture
- Data dictionaries extended by **ontologies**
- Integration of:
  - Engineering repositories and applications
  - Document repositories and applications
  - E-mail archives
  - Financial repositories and applications
  - Multimedia assets
  - Knowledge repositories and applications
- Searchable via Web Services model
Building Blocks

• Creating Knowledge Maps/Ontologies for Projects and Disciplines
  - Faceted views of the information space large enough to become a formal representation of a knowledge domain
  - Faceted by:
    • Role, discipline, project life cycle, process, organizational requirements, more

• Created by:
  - User task analysis
  - Customer observation
  - Mental modeling of work processes
Upwards Integration

- JPL Knowledge Maps to be interoperable with larger NASA taxonomy
- JPL Knowledge Maps and NASA taxonomy interoperable with larger federal taxonomies and data reference models
- JPL Knowledge Maps and NASA taxonomy interoperable with commercial taxonomies, schema and data reference models
Eventual Outcomes

- We all speak the same language
- Taxonomy work results in machine addressable schema that enable cross-application transactions
- Improved odds of successful mission outcomes by leveraging JPL's knowledge for better decision making and trade studies
Taxonomy Development
With NASA


- Case Study
- Methodology
- Best Practices
- Examples
Questions and Discussion

Thank You For Your Time!

Jayne.E.Dutra@jpl.nasa.gov