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

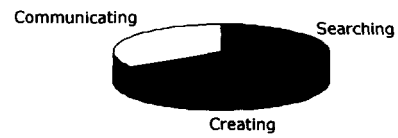
- Workers need unified, universal access to all information, but the real value comes from that portion of the information that actually solves the information problem at hand.

→ 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

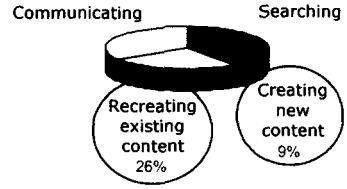
Workers spend up to 2.5 hours a day looking for information



... 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



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

The Role of Metadata

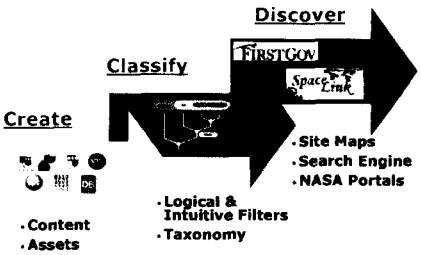
Content selection and reuse utilizes associated metadata

	Dublin Core elements *
Asset metadata—*The Who, Where and When*	
Subject metadata—*The What and Why*	
Relational metadata—*Links between Assets*	
Use metadata—*How to Monetize Assets*	

http://dublincore.org/documents/dces/

JPL Knowledge Management/ Taxonomy Strategies

Purpose of the NASA Taxonomy



Finding the right information at the right time to solve the problem at hand

JPL Knowledge Management/ Taxonomy Strategies

NASA Taxonomy Basics

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**

JPL Knowledge Management/ Taxonomy Strategies

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 (DISA)
  - Complies with metadata standards like Z39.19, ISO 2709, and Dublin Core
- Practices increase interoperability and extensibility

JPL Knowledge Management/ Taxonomy Strategies

NASA Taxonomy Top Level

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

http://eis.jpl.nasa.gov/~jsdutra/NASA_Taxonomy_ver3/Index.htm

JPL Knowledge Management/ Taxonomy Strategies

NASA Taxonomy detail

- **Disciplines**
  - **Science**
    - Aeronautics
    - Computer Sciences
    - Engineering
    - Life Sciences
    - Mathematics
    - Natural Sciences
    - Space Sciences
  - **Social Sciences**
    - Behavioral Sciences
    - Business
    - Communications
    - Economics
    - Education
    - Human factors
    - Industrial relations
    - Information Science
    - Law
    - Planning

...

- **Disciplines**
  - **Science**
    - Aeronautics
    - Computer Sciences
      - Hardware
      - Computer systems organization
      - Software
      - Data
      - Theory of computation
      - Mathematics of computing
      - Information systems
      - Computing methodologies
      - Computing applications
      - Computing milieu
    - Engineering
      - Aerospace engineering
      - Electrical engineering
      - Mechanical engineering
      - Bioengineering
    - Life Sciences
      - Agriculture
      - Biology

...

JPL Knowledge Management/ Taxonomy Strategies

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

* XML Schema (2000)

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.

NASA Taxonomy – Dublin Core Map (Draft)

Dublin Core Elements	Definition	NASA Taxonomy Mapping
Creator	Content maker.	dc:creator dc:creator.employee dc:creator.organization
Subject	Content topic.	dc:subject.organization dc:subject.missionsProjects dc:subject.disciplines
Publisher	Publisher of this manifestation.	dc:publisher.organization
Contributor	Content contributor.	dc:contributor dc:contributor.employee dc:contributor.organization
Type	Genre.	dc:type.information
Coverage	Space, period, date, jurisdiction, etc.	dc:coverage.locations dc:coverage.chronology
Audience	Content audience.	dcTERM:audience
Non DC	NASA missions and projects.	nasa:missionsProjects
Non DC	Business functions.	nasa:functions
Non DC	Technical specialties.	nasa:disciplines
Non DC	Standard industry categories.	naics:industries

Taxonomy and XML - 2

- Taxonomy is designed to support a namespace, and handles for each concept in it
- Relationships between concepts are specified by using namespace handle combinations

```
<SVTerm UID="NASA::1033079039069842">
  <label xml:lang="en">Purposes</label>
  <definition xml:lang="en">The purpose of classifying
      this learning object. LOM IEEE 1484.12.1-
      2002</definition>
  <parent UREF="NASA::1031090120256850" />
  <child UREF="NASA::1033079070955810" />
  <child UREF="NASA::1033079087128921" />
</SVTerm>
```


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

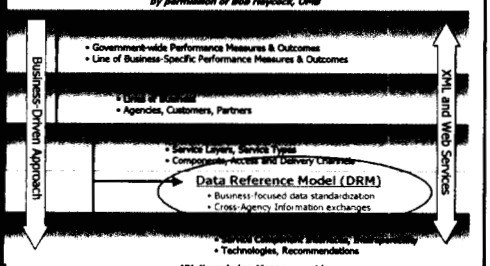
... at the Federal Level

- 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

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



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

Taxonomy Next Steps - 2

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

White Paper

Taxonomy Development With NASA

https://partners-lib.jpl.nasa.gov/partners-lib/dscgi/ds.py/Get/File-32405/NASA_Taxonomy-Dublin_Core_Paper-042203.doc

- Case Study
- Methodology
- Best Practices
- Examples