A Practitioner’s Perspective on Taxonomy, Ontology and Findability

NASA Johnson Space Center

Sarah Berndt
JSC Taxonomist
DB Consulting
Information Technology and Multimedia Services Contract
May, 2011
This case study follows the NASA “Case of Interest” definition, meaning it illustrates the value of best practices that might otherwise seem insignificant, in order to promote effective implementation.

Product Endorsement Disclaimer
This is not product endorsement but a case study on what we have done and suggestions for the recognition and development of a semantic system.
Demonstrated Need

- Need for broad categories of information that easily map to/integrate with existing information architecture
- Need for evolving procedures and processes that are easily maintained and fit users needs
- Internal & External analysis
  - Subject Matter Expert (SME) interviews
  - Repository owners, content creators
  - NASA Thesaurus
  - Dow Jones Taxonomy Services
  - Taxonomy consultants
JSC needed an uncomplicated, functional system to improve information accessibility and retrieval.


www.tatapeoplescar.com
The vision for the JSC Taxonomy is to create a controlled vocabulary to connect information stovepipes into an integrated view.

The envisioned end results are:
- Increased information accessibility, relevancy and currency
- Improvement of the information consumers’ user experiences

The scope of the JSC Taxonomy is:
- To encompass JSC created and/or owned content
- To include local level vocabularies, not to replace existing information architecture
Monetized Need

Cost Benefit Analysis

- Evidence-based measurement of JSC information workers’ search habits, contrasted with industry standards and considering the current search environment/available applications

- Information worker = content owners and creators, information consumers

- Conservative, salary based value only. Additional costs such as benefits and multiple employee efforts not included.
  
  - 8.8 hrs/wk average information worker*
  
  - 10.5 hrs/wk average JSC information worker

  $21,840 annual cost per employee

Investment that Pays

The advantages of planning and technology.

http://www.jaguarxf.info/
System Review

- International Standards Organization, ISO 2788 (establishment of a monolingual thesauri) to 25964- parts I-IV (thesauri & interoperability with other vocabularies).
- Dublin Core Metadata schemas and vocabulary type
- Z39.50- semantic protocol for search and retrieval from remote computer databases.
The Advantages of Planning and Technology

Space Shuttle Discovery, STS-133, on its final ascent, February 24, 2011

- Speeds of over 17,000 mph in ~8 minutes = acceleration of 2,000 mph each minute
- Two solid rocket boosters
- Three Space Shuttle Main Engines (SSMEs)
- Two Orbital Maneuvering System (OMS) engines to place the vehicle in orbit
- 38 primary and six vernier Reaction Control System engines for separation and in-space propulsion
Components of the Semantic System

- Controlled Vocabulary
  - Hierarchy
  - Preferred terms
- Ontology
- Equivalence Relationships
  - NonPreferred Terms

4/21/2011 JSC Chief Knowledge Officer & Information Resources Directorate
Presenter: S. Berndt, ITAMS
Components of the Semantic System (2)

Preferred terms to generate rulebases
Components of the Semantic System (3)

Rulebases are informed by the taxonomy and ontology, the proximity and location of terms, and different weights to enhance the accuracy of Classification.
Classification Driven Results

Subset of the content corpus

Fewer, more relevant results
Components of the Semantic System (4)

- Term metadata library
- Mapping to the interface
- Standards for inclusion
Term Information Driven Interface
Evolution of the Semantic System

Since January 2010, the JSC Taxonomy has more than doubled in term count while the scope of areas covered has grown exponentially!

Seven ‘Top Level’ facets are active in the JSC Google Search.

- Within the facets are over 100 ‘Classes’ (major categorizations), many of which are taxonomies in their own right.
- Within the classes are over 8,000 preferred terms with nearly 17,000 relationships between them. These relationships make up the JSC Ontology!
- Best bet URLs, images and definitions have been added to further enrich results in the Google Search Appliance.

Production Environment: JSC Taxonomy Term Growth

RTs = Related Terms
PTs = Preferred Terms
ETs = Equivalence Terms
Encourage User Participation

- Allow for user and Subject Matter Expert participation in the development and maintenance of the semantic system.
- Identifying the components and functions of a semantic system makes development more palatable. Spreadsheets are the most common tools!
Encourage User Participation

➢ Consider this a monitored folksonomy, meaning the product is centrally broadcast, not individually applied.
What We Have Learned

• Read the manual
• Examine the rulebases
• Maintain separate, but integrated domains
• Governance is both overarching and local
• Don’t require SMEs to manipulate the semantic system, but allow user participation
• Spread the word
• Test. Test. Test.
What We Suggest

• Plan your semantic system based on end user expectations and how the components of your system will meet them

• Define the scope for each component and personnel working them

• Consider the content to be represented, are there document management procedures in place to serve as a reference?

• Identify the time frame to be included: historical information retrieval, roadmap for the future or both?

• Recognize continuous maintenance and governance needs

• Inform the funders, set realistic expectations
Conclusion

At NASA Johnson Space Center (JSC), the Chief Knowledge Officer has been the champion for developing the JSC Taxonomy to capitalize on the accomplishments of yesterday while maintaining the flexibility needed for the evolving information environment of today.

A clear vision and scope for the semantic system is integral to its success. The vision for the JSC Taxonomy is to connect information stovepipes to present a unified view for information and knowledge across the Center, across organizations, and across decades.

Semantic search at JSC means seamless integration of disparate information sets into a single interface. Ever increasing use, interest, and organizational participation mark successful integration and provide the framework for future application.
Perspective has its Advantages Too!

Tracy Caldwell Dyson
ISS Expedition 24