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

Case Study: Semantic System

- 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

Functional

http://www.autoblog.com/media/2009/09/tata-nano-lux-white-630.jpg

JSC needed an uncomplicated, functional system to improve information accessibility and retrieval.

Succinct Planning = Successful Implementation

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

* International Data Corporation Hidden Costs of Information Work: A Progress Report, May 2009, Doc # 217936

Investment that Pays

http://www.jaguarxf.info/

The advantages of planning and technology.

Semantic Systems as Tools

System Review

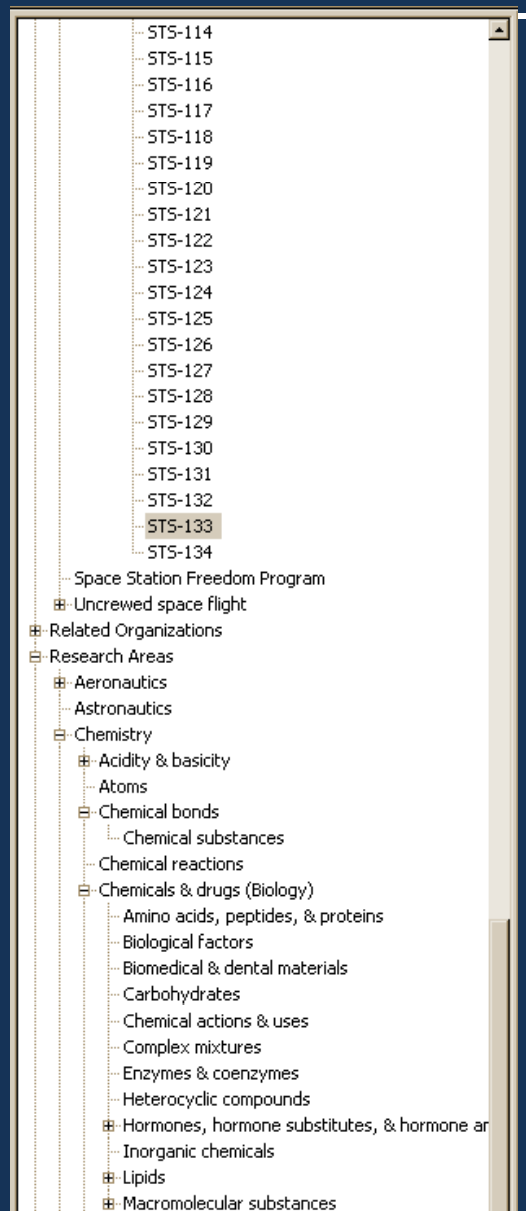
- International Standards Organization, ISO 2788 (establishment of a monolingual thesauri) to 25964- parts I-IV (thesauri & interoperability with other vocabularies).
- Dublin Core Metadata schemas and vocabulary type
- Z39.50- semantic protocol for search and retrieval from remote computer databases.

The Advantages of Planning and Technology

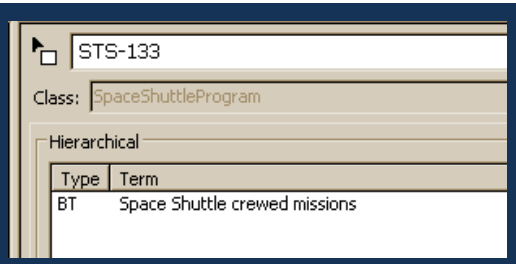
Space Shuttle Discovery, STS-133, on it's final ascent, Febrary 24, 2011

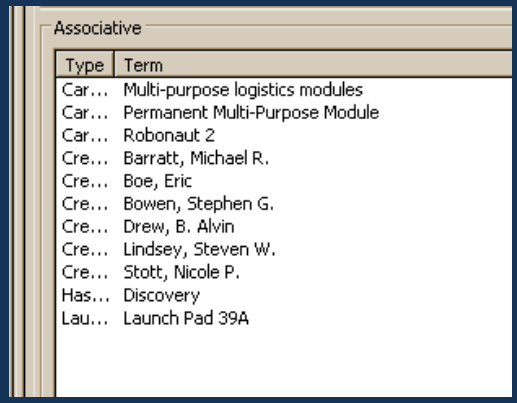
- Speeds of over 17,000 mph in ~ 8 minutes = acceleration of 2,000 mph each minute
- Two solid rocket boosters
- Three Space Shuttle Main Engines (SSMEs)
- Two Orbital Maneuvering System (OMS) engines to place the vehicle in orbit
- 38 primary and six vernier Reaction Control System engines for separation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

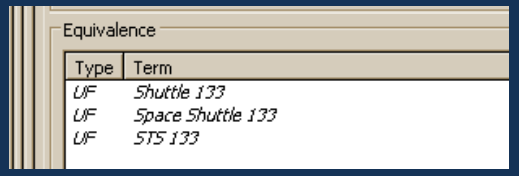
Components of the Semantic System



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







Components of the Semantic System (2)

Preferred terms to generate rulebases

STS-113_791_Missions.xml	XML Docum...	3/10/2011 11:27 AM	24,461
STS-114_792_Missions.xml	XML Docum...	3/10/2011 11:27 AM	30,100
STS-115_793_Missions.xml	XML Docum...	3/10/2011 11:27 AM	24,384
STS-116_794_Missions.xml	XML Docum...	3/10/2011 11:26 AM	24,268
STS-117_795_Missions.xml	XML Docum...	3/10/2011 11:27 AM	24,360
STS-118_796_Missions.xml	XML Docum...	3/10/2011 11:26 AM	24,459
STS-119_797_Missions.xml	XML Docum...	3/10/2011 11:28 AM	24,411
STS-120_798_Missions.xml	XML Docum...	3/10/2011 11:26 AM	24,351
STS-121_799_Missions.xml	XML Docum...	3/10/2011 11:28 AM	24,176
STS-122_800_Missions.xml	XML Docum...	3/10/2011 11:26 AM	24,266
STS-123_801_Missions.xml	XML Docum...	3/10/2011 11:26 AM	24,272
STS-124_802_Missions.xml	XML Docum...	3/10/2011 11:27 AM	24,256
STS-125_2541_Missions.xml	XML Docum...	3/10/2011 11:27 AM	41,755
STS-126_803_Missions.xml	XML Docum...	3/10/2011 11:26 AM	24,185
STS-127_23321_Missions.xml	XML Docum...	3/10/2011 11:27 AM	24,209
STS-128_23611_Missions.xml	XML Docum...	3/10/2011 11:26 AM	24,204
STS-129_23613_Missions.xml	XML Docum...	3/10/2011 11:27 AM	24,120
STS-130_23618_Missions.xml	XML Docum...	3/10/2011 11:27 AM	24,109
STS-131_23620_Missions.xml	XML Docum...	3/10/2011 11:26 AM	24,320
STS-132_31919_Missions.xml	XML Docum...	3/10/2011 11:28 AM	24,112
STS-133_31921_Missions.xml	XML Docum...	3/10/2011 11:26 AM	24,198
STS-134_31923_Missions.xml	XML Docum...	3/10/2011 11:27 AM	23,920
Stucture & Function Related Neuronal Plasticity_52631_Equipment.xml	XML Docum...	3/10/2011 11:36 AM	12,053
Stucture & Solidification of Largely Undercooled Melts_61253_Equipment.xml	XML Docum...	3/10/2011 11:36 AM	11,658
Student Temporary Employment Program_57111_Functions.xml	XML Docum...	3/10/2011 11:23 AM	30,536
Studies of Ionization States_47768_Equipment.xml	XML Docum...	3/10/2011 11:36 AM	13,963
Studio photography_32941_Functions.xml	XML Docum...	3/10/2011 11:23 AM	9,603
Study Capillary Waves on Water Surface_47502_Equipment.xml	XML Docum...	3/10/2011 11:36 AM	4,846
Study of Meteoroid Impact Craters on Various Materials_49555_Equipment.xml	XML Docum...	3/10/2011 11:36 AM	5,567
Study of the Effects of Mammalian Gravity Receptors_50175_Equipment.xml	XML Docum...	3/10/2011 11:36 AM	14,494
Study Proposed Propellant Acquisition System_47494_Equipment.xml	XML Docum...	3/10/2011 11:36 AM	3,754
Study Solar Flux Separation_47500_Equipment.xml	XML Docum...	3/10/2011 11:36 AM	3,793

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.

```
<!-- TITLE RULES FOR NPTs -->
- <combine weight="100" label="link.Missions.STS-133.31921_TITLE_NPT" _key="k815323">
  <!-- multi-word NPT - title -->
  - <max not="0" scale="0" weight="100" _key="k815325">
    - <phrase case="0" field="title" foreach="0" weight="25" not="0" stem="1" _key="k815326">
      - <any case="0" not="0" weight="100" stem="1" _key="k815327">
          <text case="0" field="title" weight="100" not="0" data="S T S" stem="1" _key="k815328" />
          <text case="0" field="title" weight="100" not="0" data="S.T.S." stem="1" _key="k815329" />
          <text case="0" field="title" weight="100" not="0" data="STS" stem="1" _key="k815330" />
        </any>
        <text case="0" field="title" weight="100" not="0" data="133" stem="1" _key="k815331" />
      </phrase>
    - <near case="0" count="2" foreach="0" weight="20" field="title" not="0" stem="1" _key="k815332">
      - <any case="0" not="0" weight="100" stem="1" _key="k815333">
          <text case="0" field="title" weight="100" not="0" data="S T S" stem="1" _key="k815334" />
          <text case="0" field="title" weight="100" not="0" data="S.T.S." stem="1" _key="k815335" />
          <text case="0" field="title" weight="100" not="0" data="STS" stem="1" _key="k815336" />
        </any>
        <text case="0" field="title" weight="100" not="0" data="133" stem="1" _key="k815337" />
      </near>
    - <sentence case="0" field="title" not="0" weight="15" stem="1" _key="k815338">
      - <any case="0" not="0" weight="100" stem="1" _key="k815339">
          <text case="0" field="title" weight="100" not="0" data="S T S" stem="1" _key="k815340" />
          <text case="0" field="title" weight="100" not="0" data="S.T.S." stem="1" _key="k815341" />
          <text case="0" field="title" weight="100" not="0" data="STS" stem="1" _key="k815342" />
        </any>
        <text case="0" field="title" weight="100" not="0" data="133" stem="1" _key="k815343" />
      </sentence>
    </max>
  - <max not="0" scale="0" weight="100" _key="k815344">
    - <phrase case="0" field="title" foreach="0" weight="25" not="0" stem="1" _key="k815345">
      - <any not="1" _key="k815346">
          <text data="Space" _key="k815347" />
        </any>
        <text case="0" field="title" weight="100" not="0" data="Shuttle" stem="1" _key="k815348" />
        <text case="0" field="title" weight="100" not="0" data="133" stem="1" _key="k815349" />
      </phrase>
```

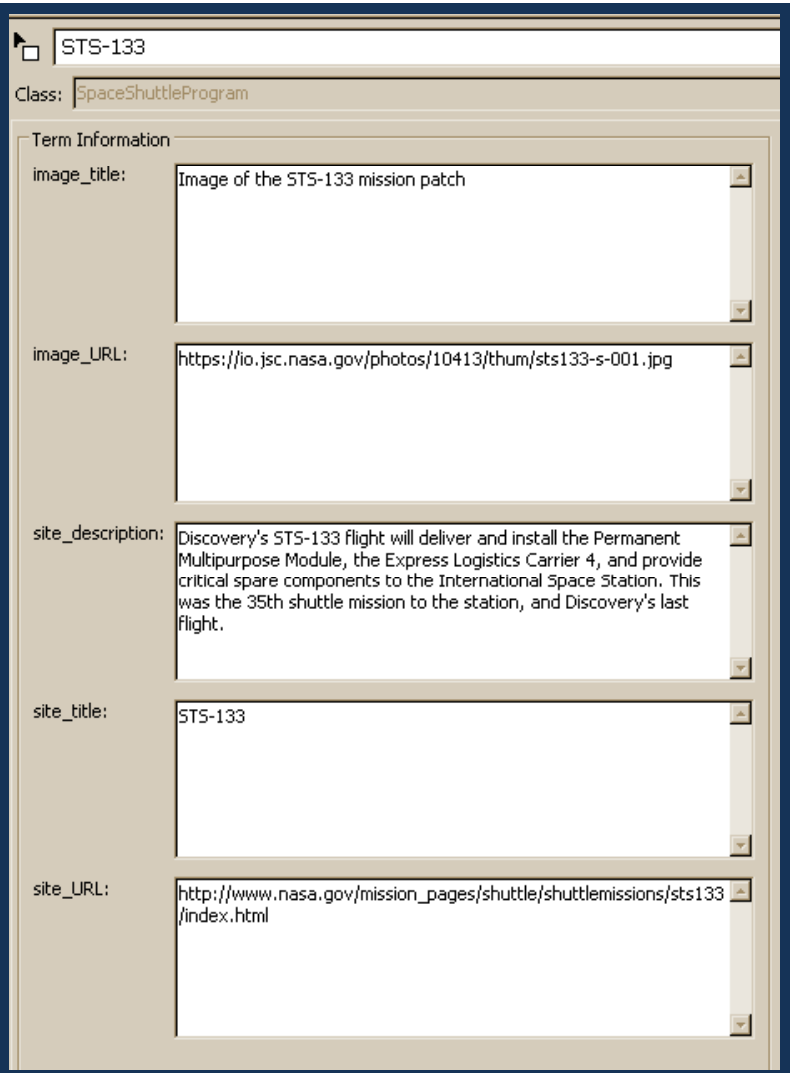
Classification Driven Results

Subset of the content corpus

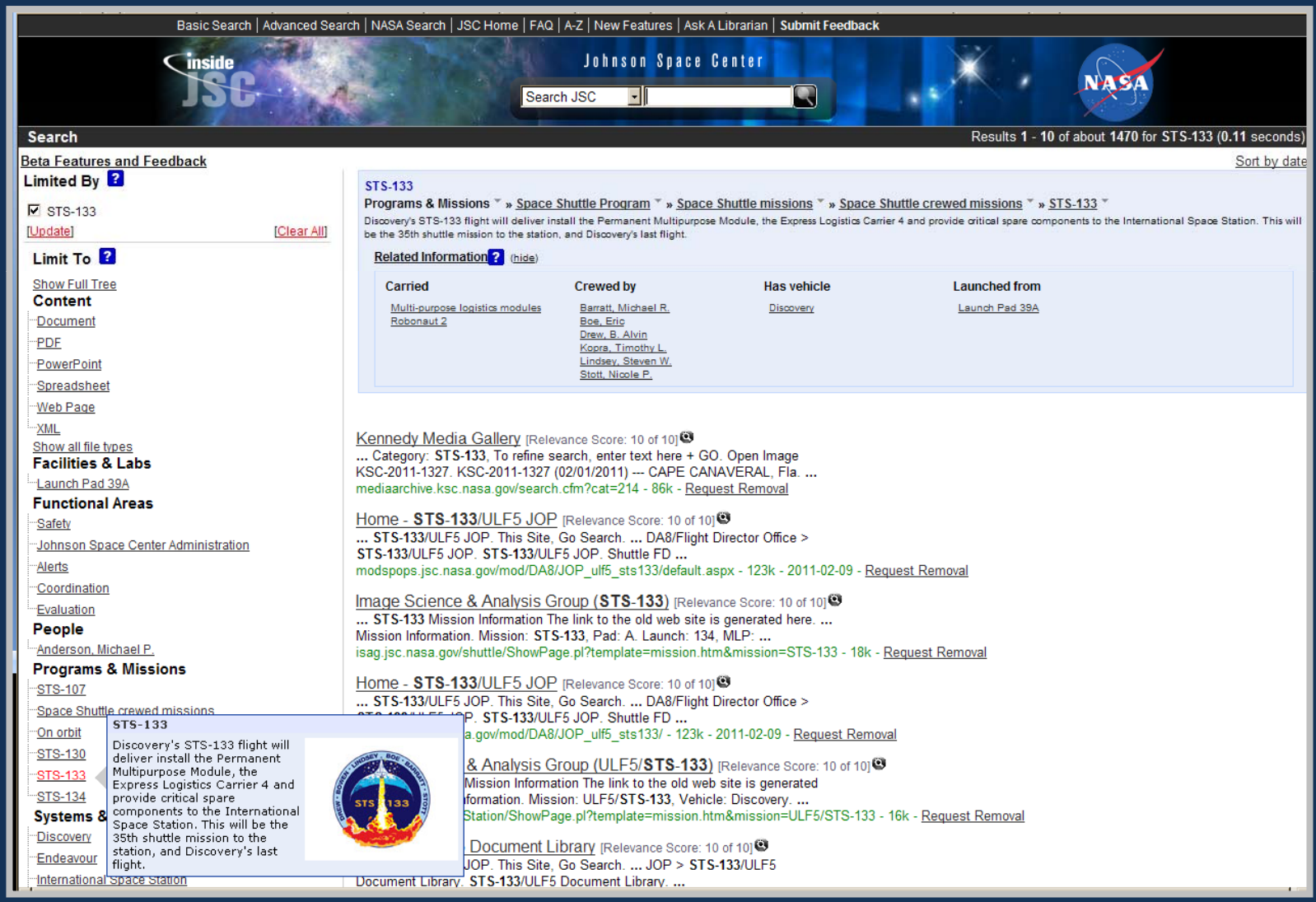
Fewer, more relevant results

Components of the Semantic System (4)

- Term metadata library
- Mapping to the interface
- Standards for inclusion



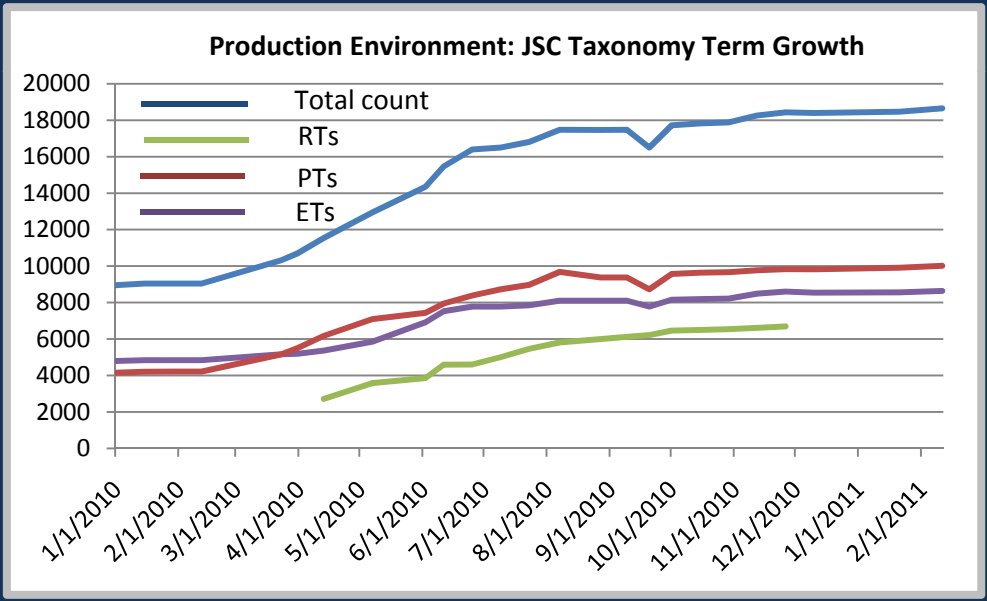
Term Information Driven Interface



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



RTs= Related Terms
PTs= Preferred Terms
ETs= Equivalence Terms

Encourage User Participation

A	B	C	D	E	F	G	H	I	J	K	L	M	
Preferred Term										**Level**		**Scope Notes**	**Comments**
				STS-126						(level 4)			
				STS-127						(level 4)			
				STS-128						(level 4)			
				STS-129						(level 4)			
				STS-130						(level 4)			
				STS-131						(level 4)			
				STS-132						(level 4)			
				STS-133						(level 4)			
				STS-134						(level 4)			
	Space Station Freedom Program									(level 1)			
	Uncrewed space flight									(level 1)			
		Advanced Composition Explorer Mission								(level 2)		An Explorer mission managed by the Office of Space Science Mission and Payload Development Division focused on the study of accelerated particles arriving from the Sun, interstellar, and galactic sources.	
		Advanced Satellite for Cosmology & Astrop								(level 2)			
		Aquarius Mission								(level 2)			
		Asteroidal Cometary Missions								(level 2)			
			Comet Nucleus Tour Mission							(level 3)			

- 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

apod.nasa.gov/apod/ap101115.html