



Pushing the Envelope at NASA

The Development and Use of
Large Scale Taxonomies

*Jayne Dutra
Delphi Information Intelligence
Jet Propulsion Laboratory
California Institute of Technology*

April 20, 2005



Parts Catalogues



Engineering
Repositories

Electronic
Libraries

Where do I find it?



E-Mail Archives



Financial Data



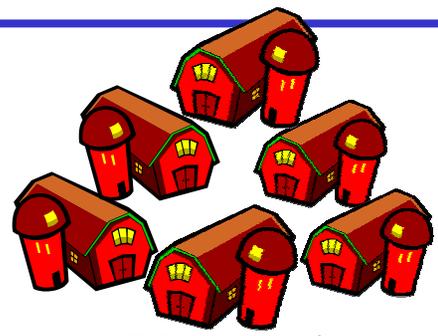
Problem
Reporting



NASA Today



JPL



Kennedy



Johnson



Langley



Ames



Goddard



Marshall

Does Your Organization Look Like This?



Chances of Finding Needed Information in a Timely Fashion

0%



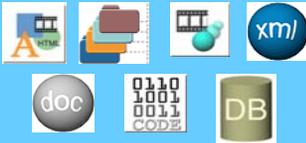
- Standards are helpful!
- Make it easy for various audiences to find relevant information
 - Provide quick access for NASA Web resources
 - Share knowledge by enabling users to easily find text files, databases and tools
 - Provide search results targeted to user interests
 - Enable the ability to move content through the enterprise to where it is needed most
 - Facilitate Records Retention and Management
- Comply with the eGov Act of 2002



Life Cycle of Electronic Content in the Real Time Organization

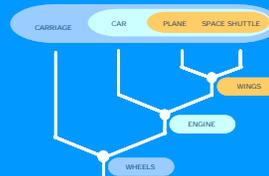


Create



- Content
- Assets

Classify



- Logical & Intuitive Filters

• **Taxonomy**

Discover



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

- Site Maps
- Search Engines
- NASA Portals
- Content Integration Networks

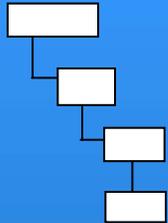
Finding the right information at the right time to solve the problem at hand



So How Did We Start?



- **Content Audit and Analysis**
 - Where is it? How is it published?
 - What is it about? Who cares?
- **Communities Definition**
 - What are the significant knowledge domains?
 - Stakeholders, publishers, consumers
- **Semantic Frameworks**
 - Controlled vocabularies from subject matter experts
 - Conversations with "rocket scientists" (and managers and engineers and accountants, etc, etc)



What is the NASA Taxonomy?

- A classification scheme meant to encompass all of NASA web content, 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
 - **A *facet* is a branch of the taxonomy**
 - **Not all facets need to be used in each instance**
 - **A facet is repeatable**
 - **The taxonomy is modular and dynamic**



*Best Practices increase interoperability
and extensibility*

- **Faceted Classification Schema**
 - Facets give flexibility and power
 - Modular in nature for easier maintenance
 - Can tag what is appropriate to the use case
- **Polyhierarchy**
 - Concepts can appear more than once
 - Enables knowledge discovery from multiple viewpoints
 - User-centric organization



- Hierarchical Granularity
 - Different levels of depth depending on attribute set and content
 - The NASA taxonomy is broad in nature by design
 - Integration points allow for mapping of local vocabulary terms back to larger semantic framework
 - Enables schema reconciliation
- Use of Existing Standards
 - Incorporates existing federal and industry terminology standards like NASA AFS, NASA CMS, FEA BRM, NAICS, and IEEE LOM
 - Provides for NASA XML namespace registry (DISA)
 - Complies with metadata standards like Z39.19, ISO 2709, and Dublin Core



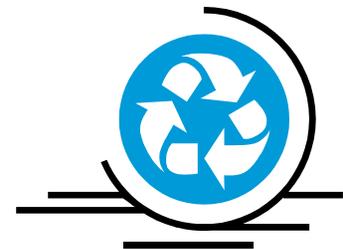
But Is It Right??



Test and Validation Phase

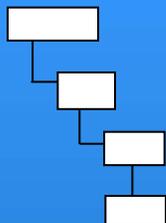
- Qualitative validation
 - Confirm stakeholders and communities
 - 71 interviews completed
- Quantitative validation
 - Select and build test collection
 - Stratify automated categorizer
 - Taxonomy test and demonstration - Seamark
- Extend taxonomy value space as needed

This work is iterative





Expressing Strategic Value With Semantic Frameworks



Facets	Strategic Value
Access Requirements	Sensitivity and access control
Audiences	Who is the content intended for
Business Purpose	Why the content was created
Competencies	Relevant field or discipline
Content Types	The genre of the content
Industries	External partners & businesses
Instruments	Flight payloads that yield science
Locations	Sites where work occurs – on and off Earth
Missions/Projects	NASA's lines of business
Organizations	NASA organizations
Subject Categories	The topic of the content



NASA Taxonomy Website



The screenshot shows the NASA Taxonomy website interface. At the top left is the NASA logo and the text "NATIONAL AERONAUTICS AND SPACE ADMINISTRATION". To the right are links: "+ Visit NASA.gov" and "+ Contact the NASA Curator". Below this is a navigation bar with five tabs: "- TAXONOMY TOP LEVEL FACETS" (selected), "+ FAQs", "+ NASA METADATA", "+ NASA TAXONOMY XML", and "+ NASA XML PROJECT".

The main content area is titled "NASA Taxonomy - Top Level Facets" and includes a list of links:

- [Access Security Requirements](#)
- [Audiences](#)
- [Business Purpose](#)
- [Competencies](#)
- [Content Types](#)
- [Industries](#)
- [Instruments](#)
- [Locations](#)
- [Missions and Projects](#)
- [Organizations](#)
- [Subject Categories](#)

On the right side, there is a section titled "What is the NASA taxonomy?" with a sub-heading "Tips on using the NASA Taxonomy". The text explains that the taxonomy is a controlled vocabulary designed to populate the NASA metadata core specification and is used for tagging NASA content.

Four yellow callout boxes with arrows point to specific elements:

- "Link to Metadata Specification" points to the "+ NASA METADATA" tab.
- "Link to XML DTDs and RDFs" points to the "+ NASA XML PROJECT" tab.
- "Background and training materials" points to the "Subject Categories" link in the list.
- "Links to Controlled Vocabularies" points to the "Subject Categories" link in the list.

The footer contains the "FIRST GOV" logo with the tagline "Your First Click to the U.S. Government" and links to "Freedom of Information Act", "The President's Management Agenda", and "FY2002 Agency Performance Report". To the right is the NASA logo and contact information for the NASA Curator, Jayne Dutra, with a last update date of May 25, 2004.

<http://nasataxonomy.jpl.nasa.gov>



• Content Types

- . Announcements
 - . . Press Kits
 - . . Press Releases
- . Articles, Notes, and Papers
- . Calendars and Schedules
 - . . Agendas
- . Case Studies
- . Catalogs and Databases
- . Correspondence
 - . . e-Mails
 - . . Memos
- . Databases
 - . . Bibliographic Databases
 - . . Image Databases
 - . . Designs and Specifications
 - . . Configuration Controls
 - . . Notebooks
 - . . Quality Control
 - . . Requirements
- . Drawings
- . Educational Materials
 - . . Activity Guides
 - . . Educational Toys
 - . . Educator's Guides

Configuration Controls

Broader Terms:

[Designs and Specifications](#)

Scope Note:

Records of changes to documentation or hardware, including engineering change requests and waivers.

Term Number:

52

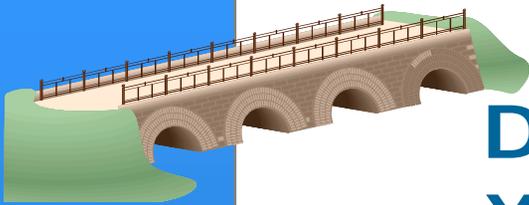


Using the Taxonomy



- 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 the DC metadata tags

Field	Name space	Definition	Data Type or Source
Type	dc:type	The nature or genre of the content of the resource	Values come from NASA taxonomy facet: Content Types



Dublin Core Mapping and XML Schema Development

- Dublin Core metadata mapping where appropriate
- Created any necessary NASA specific tags
 - Some datasets unique to the Agency
- Developed XML schema from metadata
 - RDF Files enable easy reuse for developers
- Next Steps:
 - Educate and train publishing communities



Selected and Built Test Collection



Collection	Source URL	No of Docs
Lessons Learned Database	http://llis.nasa.gov	1,370
NTRS (NASA Technical Report Server)	http://ntrs.nasa.gov	213,900
SIRTF (Space Infrared Telescope Facility) Project Library	http://sirtifweb.jpl.nasa.gov	4,054
James Webb Space Telescope (JWST) Project Documents	http://ngst.gsfc.nasa.gov/doclist/bytitle.html	634

<http://www.siderean.com/nasa/nasademo.jsp>
Demo using Seamark from Siderean Software



Built Demonstration of Taxonomy Value in Search and Navigation



NATIONAL AERONAUTICS
AND SPACE ADMINISTRATION

Search

219958 items

by Organization

- [NASA Affiliated Institutions](#) 1378
- [NASA Centers](#) 76545
- [NASA Contractors](#) 10108
- [NASA Enterprises](#) 815
- [NASA Headquarters](#) 4042
- [Other NASA Partners](#) 999

by Competencies

- [Business](#) 386
- [Engineering](#) 393
- [Mission](#) 555
- [Scientific](#) 410
- [Technical](#) 218

by Subject

- [Aeronautics](#) 26532
- [Astronautics](#) 31758
- [Chemistry and Materials](#) 17086
- [Engineering](#) 39631
- [Geosciences](#) 30770
- [Mathematical and Computer Sciences](#) 13286
- [Space Sciences](#) 22685
- [4 more](#)

by Information Type

- [Catalogs and Databases](#) 32
- [Designs and Specifications](#) 62
- [Plans and Agendas](#) 158
- [Results and Analyses](#) 260
- [Reviews and Lessons Learned](#) 1819
- [Status Reports](#) 119
- [Technical Reports](#) 229
- [6 more](#)

by Missions and Projects

- [Aerospace Technology](#) 60
- [Biological and Physical Research](#) 68
- [Data](#) 140
- [Earth Sciences](#) 1497
- [Human Exploration and Development...](#) 10680
- [Planetary Missions](#) 4819
- [Space Sciences](#) 9467

by Collection

- [Lessons Learned](#) 1370
- [NTRS](#) 213900
- [SIRTF](#) 4054
- [Webb](#) 634

by Date

- [1972](#) 8392
- [1973](#) 8512
- [1974](#) 7828
- [1975](#) 7704
- [1992](#) 8131
- [1993](#) 8519
- [1994](#) 7712
- [74 more](#)

...that provides common access framework across test collections



- Taxonomy stewardship
 - Governance
 - Maintenance, versioning
 - Education and training
 - Facilitate standard adoption process
- Apply in public and internal portals, applications and repositories
 - DDM, PDMS, CMS, other systems
 - Search integration
 - Faceted search and navigation
 - Content integration networks for real time delivery



Thanks for Your Time!



Jayne.E.Dutra@jpl.nasa.gov
California Institute of Technology

And remember:

Just say NO!





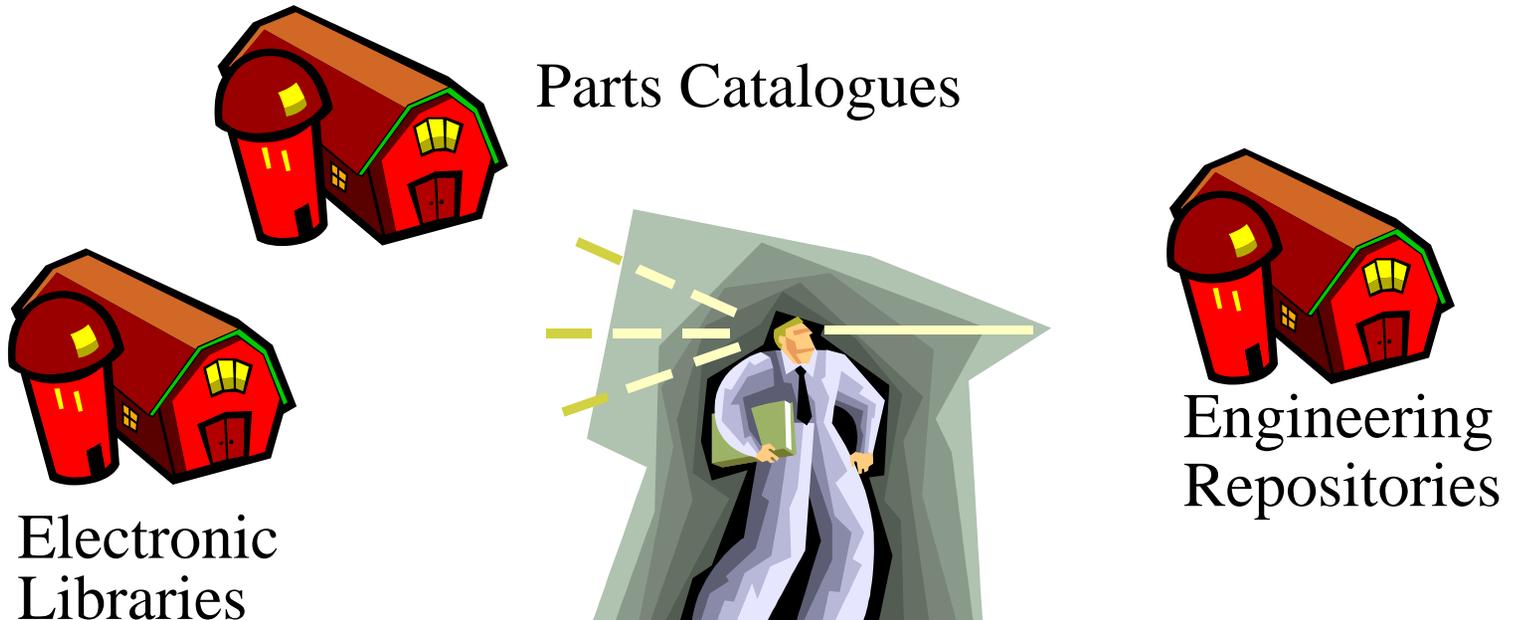
White Papers



- White Paper: Taxonomy Development With NASA, Dutra and Busch, 2003
 - https://pub-lib.jpl.nasa.gov/pub-lib/dscgi/ds.py/Get/File-22/NASA_Taxonomy_Overview-0203.ppt
- White Paper: Implementing the NASA Taxonomy Through Service Oriented Architectures, Dutra and Xiao, 2/2004
 - https://pub-lib.jpl.nasa.gov/pub-lib/dscgi/ds.py/Get/File-118/Content_Integration_Networks_WP_02_04_04.doc



*Applying a larger semantic framework
to local environments to create an
integrated, nested information
architecture*



Where do I find it?





What's Needed At The Center Level

- A **unified view** of the project information space
 - Abstracted away from any particular technology or repository
 - Able to handle “harmonization” between repositories according to a larger point of view
- Common data models within and across JPL knowledge domains
 - Common data elements (i.e. JPL Taxonomy and definitions)
 - Relationships between elements
 - Metadata mappings from local models to domain models
- Semantic models that are part of a cohesive lab wide interoperable information architecture



Case Study Goal: Allow Cassini flight project operations teams to match anomalous behavior from spacecraft to engineering design specifications for problem resolution.

1. Characterize targeted databases/repositories
Problem Failure Reporting System, Electronic Libraries, PDMS, Risk Management DB, etc.
2. Create RDF from data architectures
3. Queries identify fields of interest using semantic properties and return **integrated** result sets



Cassini Sample of Unified Search



Collections: Problem Failure Reporting System and the Cassini Electronic Library

Not a common metadata schema

PFR:

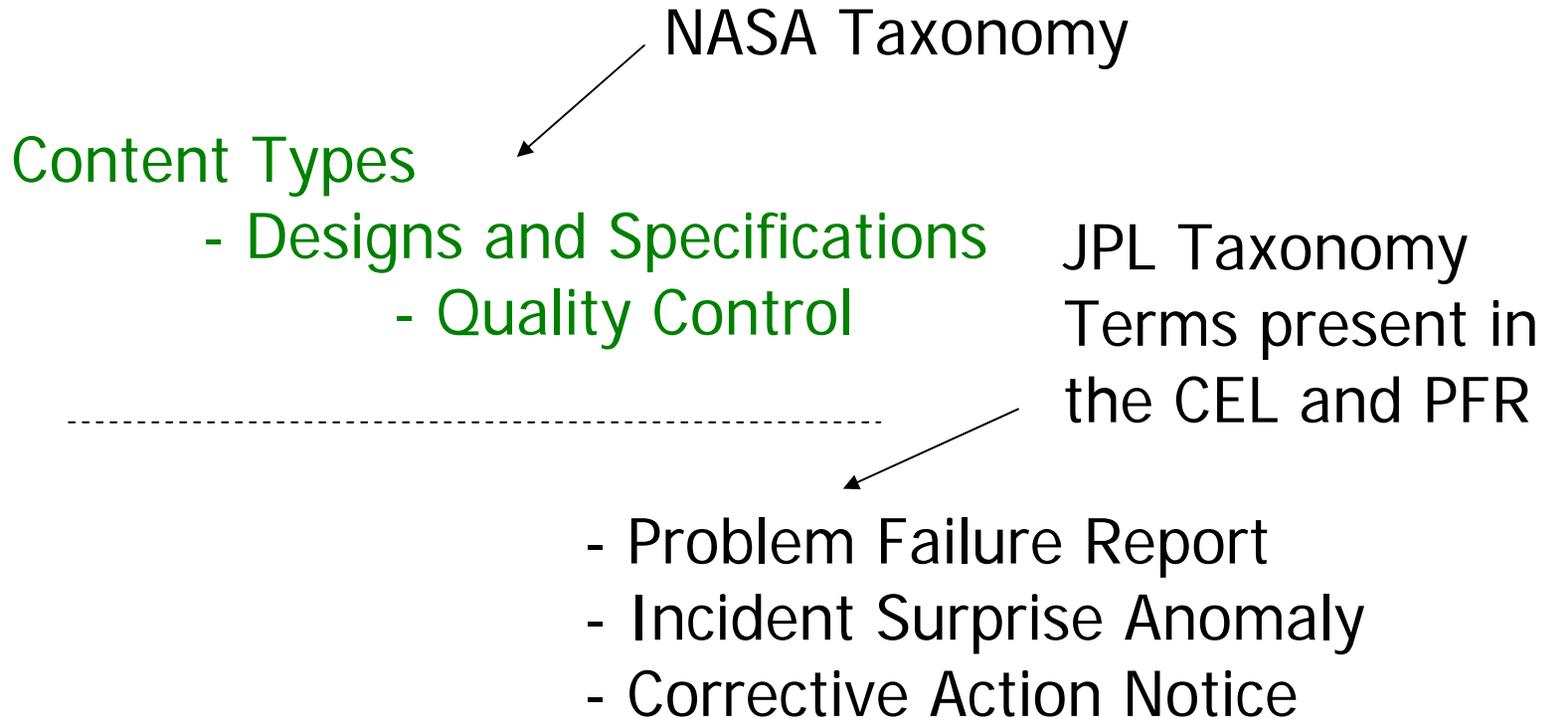
- Project Name
- Anomaly Type
- Subsystem
- Report Status
- Date

CEL:

- Project Name
- Content Type
- System
- Project level
- Responsible Team/WBS
- Date



NASA Taxonomy Transitioned to a JPL Taxonomy



Re- Combined through RDF



Connecting heterogeneous collections:
PFR System & the Cassini Electronic Library
Mapping fields to each other using semantic hierarchies.

Search and Browse the catalogue by:

- Project Name
- Content Type
- System
- Subsystem
- Responsible Team/WBS
- Date
- Collection

Next terms mapped to
each other using an existing
spacecraft ontology



Final Results of Data Harmonization



- A system whereby the user can browse all documents relating to the Cassini camera and its subsystem independent of any particular repository's search engine.
- Harmonization achieved by mapping terms to a common vocabulary (the Taxonomy)
- Could browse by:
 - System, Sub-system
 - Instrument
 - Content Type – PFRs, ECR's, Designs Specs, etc.
 - WBS or Responsible Team
 - Date



Achieving the Vision



- Leverage what projects produce in the normal course of their business
 - WBS lists
 - Document trees
 - Document matrices
 - Document standards, Flight Project Practices processes
- *There are many un-mined sources for semantic processing*
- *What schema already exist in your organization?*