



# **NASA Taxonomy: A Way to Unify the NASA Information Space**

ISKO Singapore

# KOS create order & makes sense of things



Vehrli. The art of clean up: Life made neat and tidy. (<http://www.fubiz.net/2011/08/31/the-art-of-clean-up/>)

## Purpose of KOS

| Purpose            | Description                                                                         |
|--------------------|-------------------------------------------------------------------------------------|
| <b>Translation</b> | Translate user queries into information retrieval indexing vocabulary.              |
| <b>Consistency</b> | Enable complete and consistent attribute values.                                    |
| <b>Semantics</b>   | Specify semantic relationships between and among terms.                             |
| <b>Browsing</b>    | Enable users to navigate hierarchies and browse categories to locate content items. |
| <b>Retrieval</b>   | Aid to help users think about how to search for content.                            |

After: ANSI/NISO Z39.19-2005 (r2010)

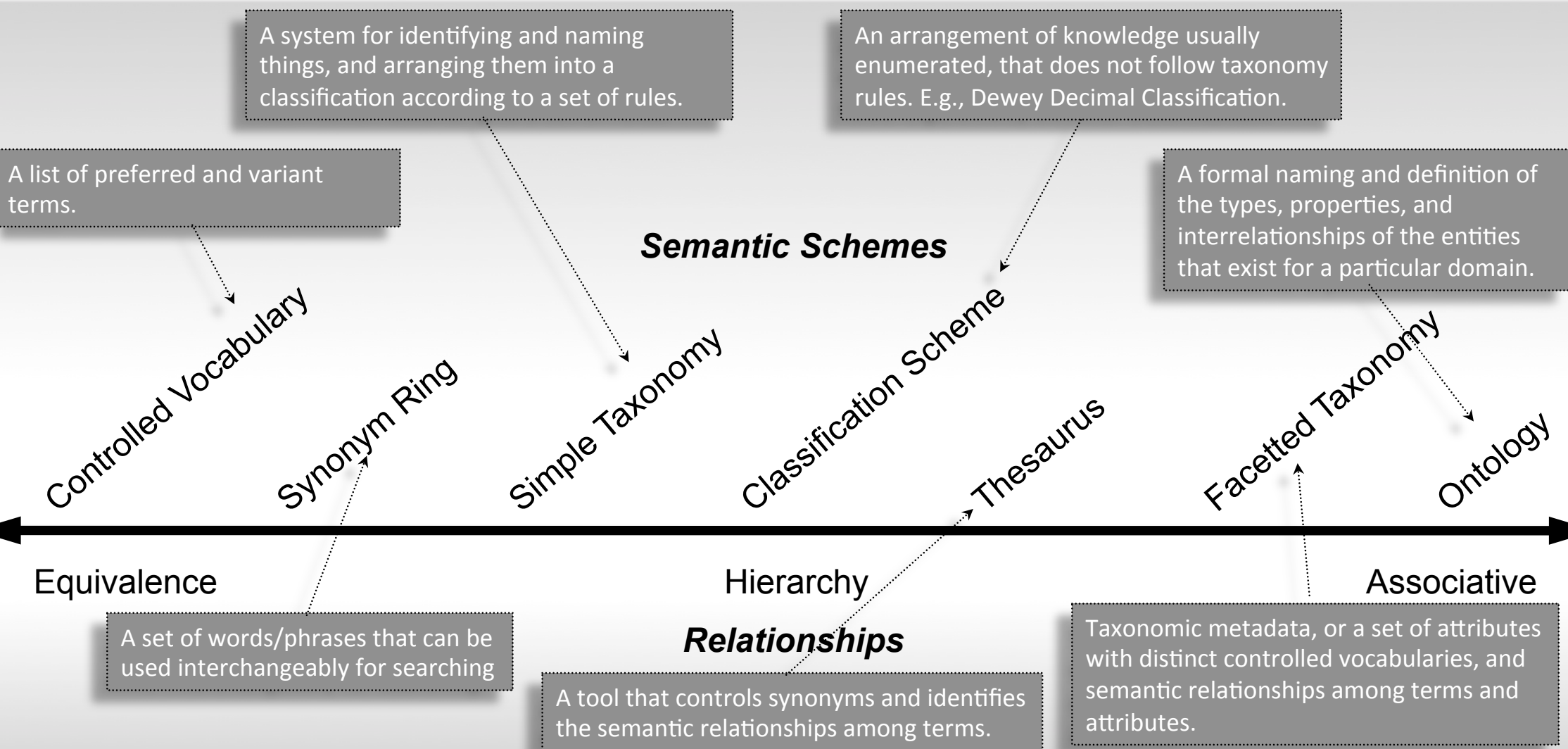
# Principles of vocabulary control

| Principle                                  | Description                                                  | Example                                        |
|--------------------------------------------|--------------------------------------------------------------|------------------------------------------------|
| <b>Eliminate ambiguity</b>                 | Ensure that each term has only one meaning                   | Drum (container) vs. Drum (musical instrument) |
| <b>Control synonyms</b>                    | Identify preferred label for each context. Concept vs. label | IBM vs. International Business Machines        |
| <b>Establish relationships among terms</b> | Equivalence, hierarchy and associative relationships         |                                                |
| <b>Test, validate and maintain terms</b>   | Query logs and content analytics                             |                                                |

## Using warrant to select terms

| Type                   | Description                                                                       |
|------------------------|-----------------------------------------------------------------------------------|
| Literary warrant       | The label that most commonly appears in publications (based on natural language). |
| Organisational warrant | The official label (based on organisational needs, priorities or policies).       |
| User warrant           | The label users most commonly use.                                                |

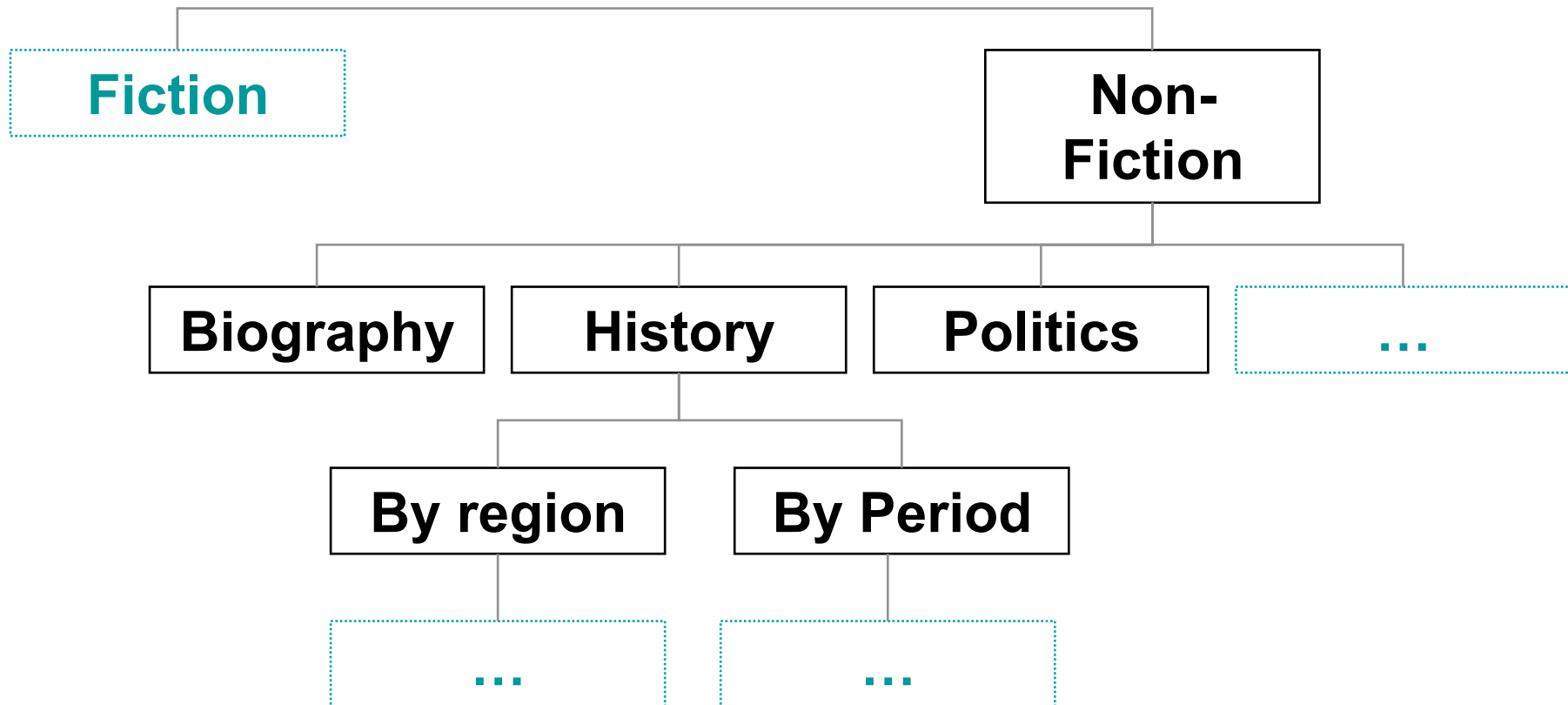
# KOS Schemes: Simple to Complex



After: Amy Warner. *Metadata and Taxonomies for a More Flexible Information Architecture*

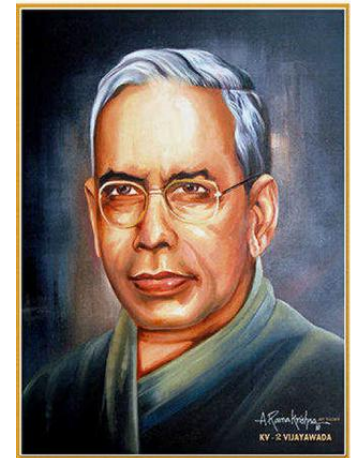
# What is a taxonomy?

- ❖ A taxonomy is a particular form of controlled vocabulary in which the labels are organised according to a hierarchy.



# Origins of faceted classification

- ❖ Mathematician/librarian S.R. Ranganathan (1920s)
- ❖ Developed as an alternative to Dewey Decimal System for books.
- ❖ “Colon Classification” facets
  - 1) Personality – topic or orientation
  - 2) Matter – things or materials
  - 3) Energy – actions
  - 4) Space – places or locations
  - 5) Time – times or time periods



S.R. Ranganathan. Painting by A. Ramakrishna, Art teacher, K.V. No. 2, Vijayawada ([http://www.thehindu.com/multimedia/dynamic/01548/12isbs-ranga\\_G4\\_12\\_1548490e.jpg](http://www.thehindu.com/multimedia/dynamic/01548/12isbs-ranga_G4_12_1548490e.jpg))

# 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.
- ❖ Taxonomic metadata.

**Facets = Metadata (with Controlled Values)**

# Facetted classification

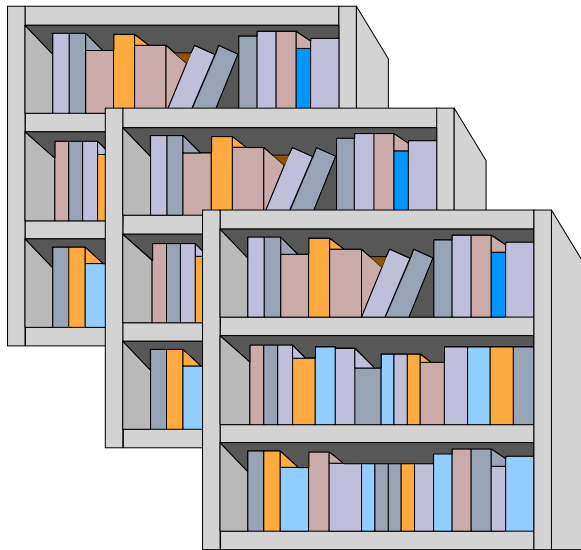
- ❖ Categorises items into multiple taxonomies or “stackonomies” based on unique but pervasive characteristics such as geography, type, price, etc.
  - Wines by region *France* > *Alsace*
  - Wines by type *White* > *Chardonnay*
  - Wines by price
- ❖ Reflects the domain of the items and allows assignment of **multiple classifications** to items
  - By region, by type, by price, etc.
- ❖ Intended for searching with multiple terms in combination (post-coordination), one from each facet.



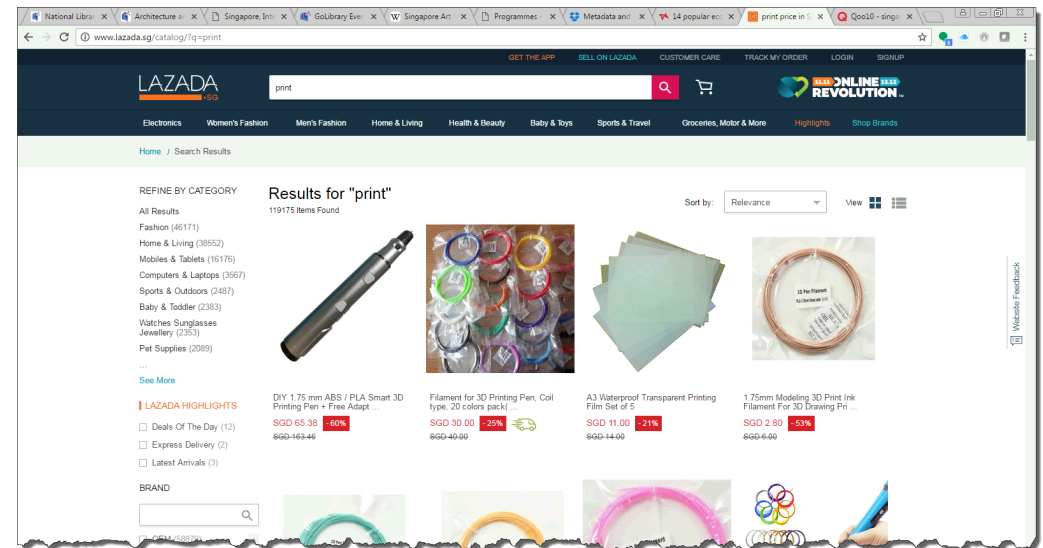
Different types of wine published by WI  
Folly <http://visual.ly/different-types-wi>

# How is a taxonomy different from a library cataloging?

- ❖ Taxonomy is a post-coordinated indexing system
- ❖ It relies on named entities (people, organisations, locations, etc.)



**Pre-coordinated system: Each book has its place on the shelf**



**Post-coordinated system: Items can be in many categories**

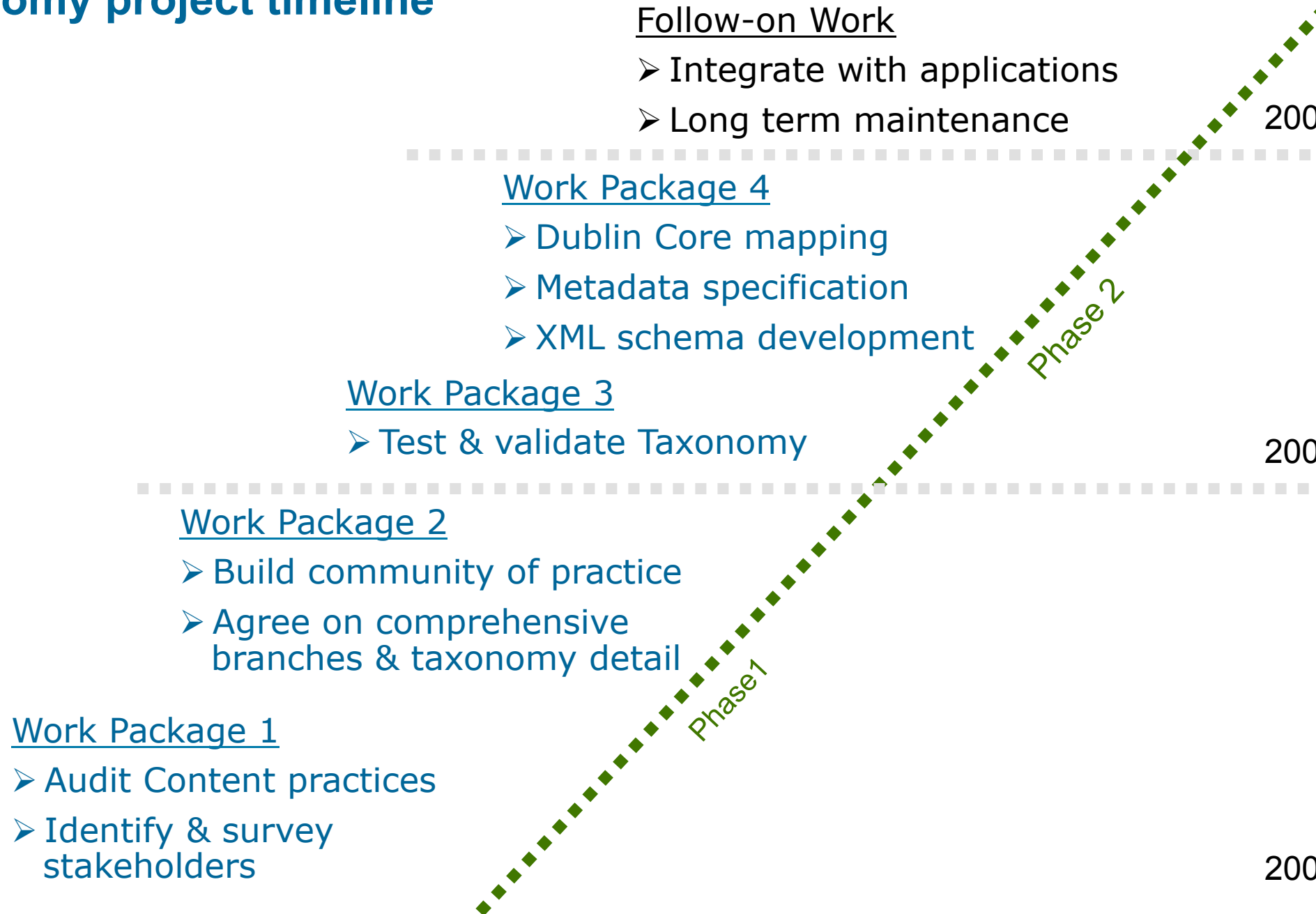


## Case study: NASA Taxonomy goals and objectives

- ❖ Make it easy for various audiences to find relevant information from NASA programs quickly
  - Provide easy access for NASA Web resources.
  - Information integration for unified queries and management reporting.
  - Improve search results targeted to user interests.
  - Enable the ability to move content through the enterprise to where it is needed most.
  - Facilitate Records Management and Retention Requirements.
- ❖ Comply with E-Government Act of 2002.
- ❖ Be ready to participate in federal XML projects.

**Taxonomy goal = Enable knowledge discovery**

# IASA Taxonomy project timeline

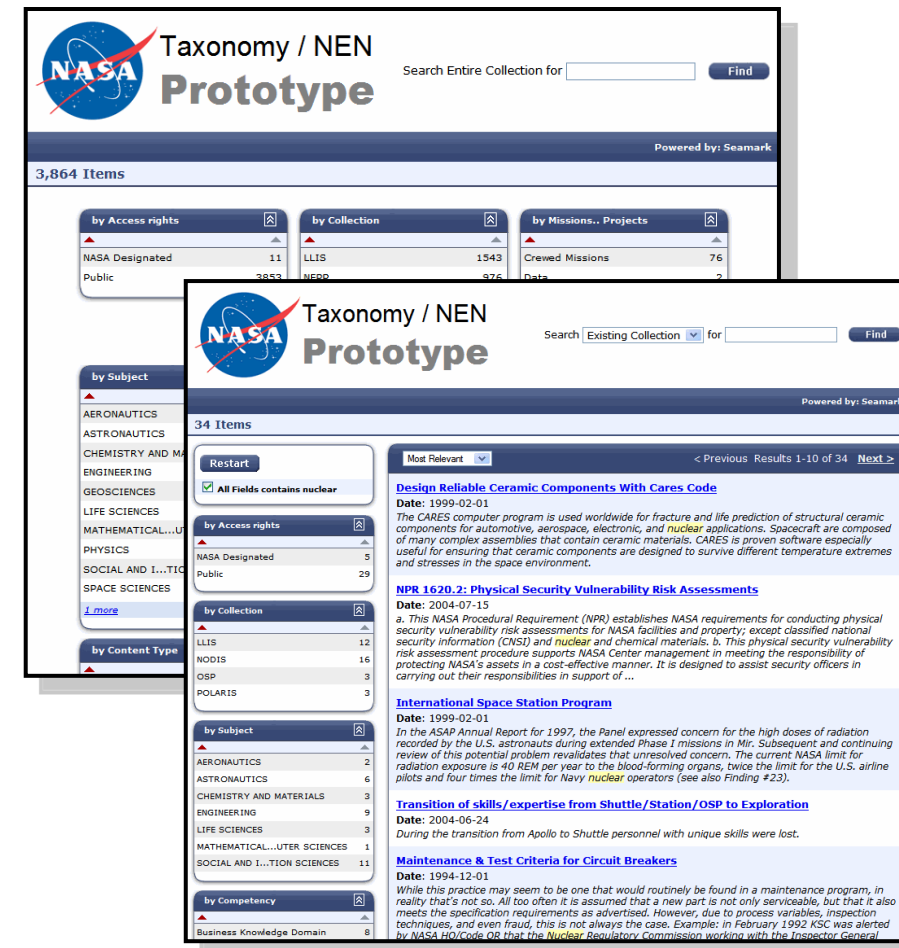


## Phase I activities (2002): Fact finding & community building

- ❖ Review existing taxonomy terms in light of:
  - Mission and Organizational changes
  - Use cases
  - New content
- ❖ Identify potential Gold Sources and subject matter experts
- ❖ Make revisions as necessary
- ❖ Meet with Review Team regularly for NASA input

# Phase 2 activities (2003): Validation & standardisation

- ❖ Application of NASA Taxonomy to repositories
  - 5 content sources – LLIS, NODIS, NEPP, OSP, and POLARIS.
  - Map existing metadata schema from test repositories to NASA Taxonomy.
  - Review outcome of faceted navigation test in web environment.
  - Change Taxonomy as needed from test results.
- ❖ Develop technical specifications
  - Dublin Core mapping.
  - Metadata specification.
  - XML schema development.



## Taxonomy follow-on work (2004+)

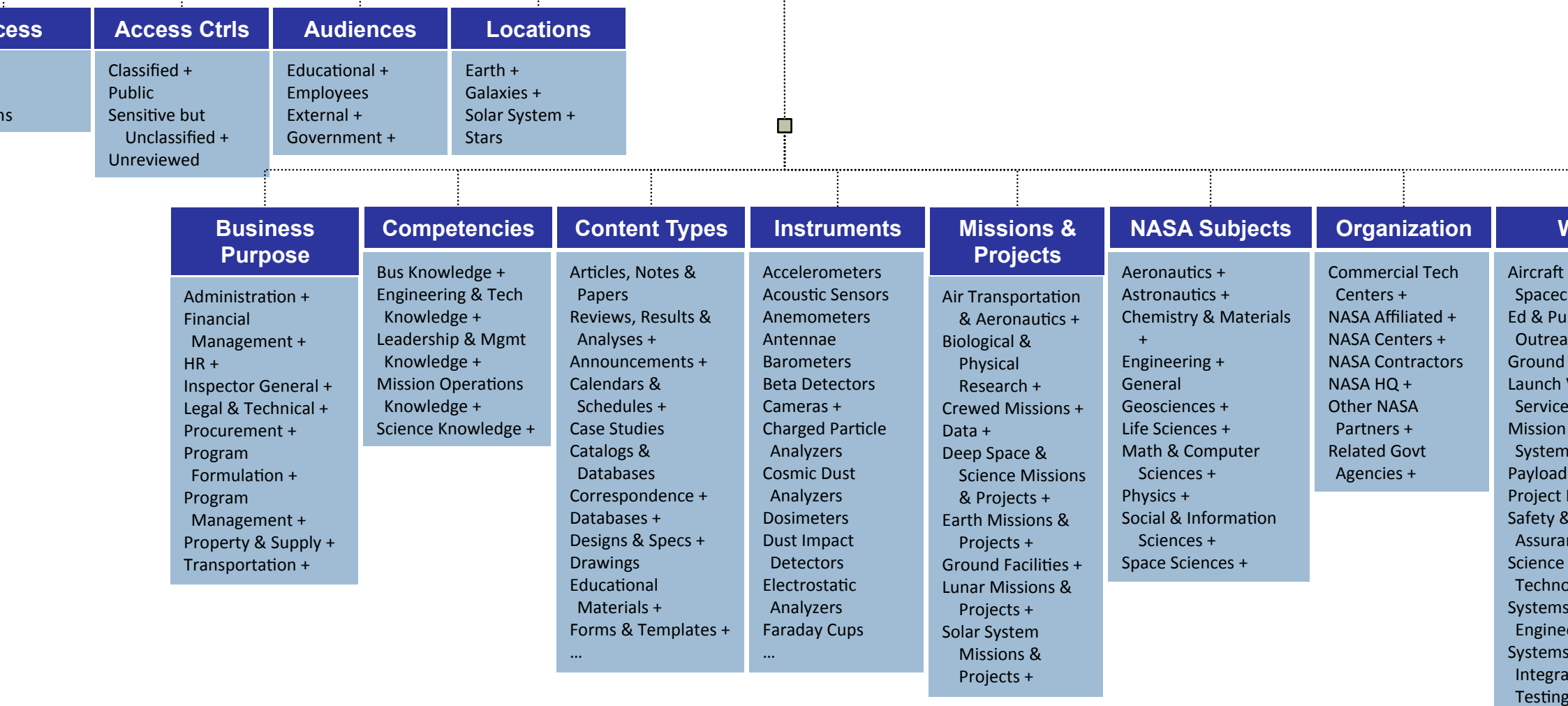
- ❖ Taxonomy stewardship
  - Maintenance, education and training
  - Facilitate standard adoption process
- ❖ Apply in public and internal portals
  - Public and Engineering portals
  - Search integration
    - Verity K2
    - Faceted search and navigation - Seamark
  - Content integration networks for real time delivery
- ❖ NASA Enterprise Architecture Group standards development
  - Agency Registries, RSS Syndication, Semantic Web components

# NASA Taxonomy facets



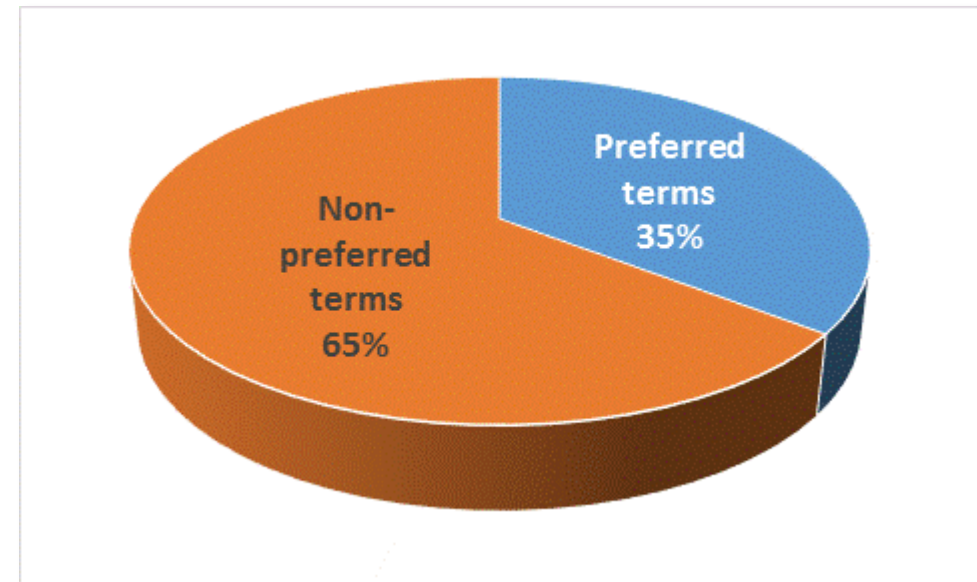
| Facets                    | Strategic Value                            |
|---------------------------|--------------------------------------------|
| <b>Access</b>             | Overall access filtering                   |
| <b>Access Controls</b>    | Sensitivity and access control             |
|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

# NASA Taxonomy top 2 levels



# IASA Taxonomy statistics

- ❖ Total number of terms: 5533
  - Preferred terms: 1961
  - Non-preferred terms: 3572
- ❖ Hierarchical relationships: 2013
  - Top terms: 12
  - Equivalence relationships: 3718
  - Associative relationships: 0



# NASA Taxonomy website

**Taxonomy  
change  
process**

**Background  
& training  
materials**

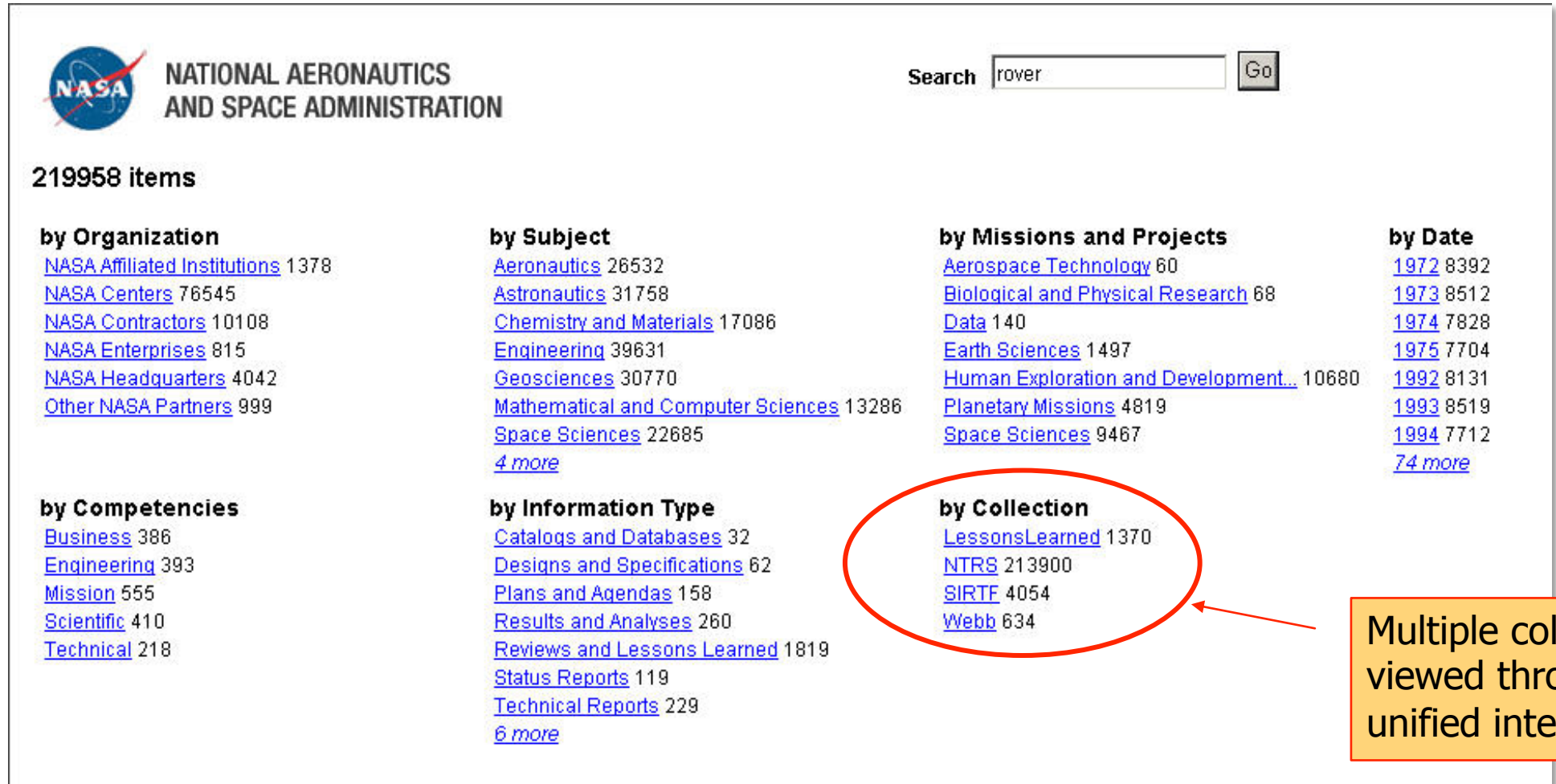
- Metadata spec
- XML RDF schema
- SKOS files
- Editorial style guide

**Links to  
current  
vocabularies**

The screenshot shows the NASA Taxonomy website. At the top 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: '- CURRENT VERSION' (highlighted in blue), '+ MAINTENANCE', '+ RESOURCES', '+ FOR DEVELOPERS', and '+ ARCHIVES'. A red arrow points from the 'Taxonomy change process' box to the 'MAINTENANCE' tab. Below the navigation bar is a large banner image with the text 'NASA TAXONOMY' in large white letters. To the left of the banner is a blue sidebar with the text 'Current Version' and two links: 'FACETS BY HIERARCHY' and 'FACETS BY A-Z INDEX'. To the right of the banner is a section titled 'FACETS BY HIERARCHY' with a paragraph: 'Taxonomies have tree structures with branches or facets. Below are the facets for the current NASA Taxonomy, version 2.0'. Below this paragraph is a list of facets: Access Control, Audiences, Business Purpose, Competencies, Content Types, Instruments, Locations, Missions and Projects, Organizations, STI Subject Categories, and Work Breakdown Structure. A red arrow points from the 'Links to current vocabularies' box to this list. At the bottom of the page is a footer with the 'FIRSTGov' logo and a list of links: '+ Freedom of Information Act', '+ The President's Management Agenda', '+ FY2002 Agency Performance Report', '+ NASA Privacy Statement, Disclaimer, and Accessibility Certification', and '+ Freedom to Manage'. To the right of this list is the NASA logo and contact information: 'Contact the NASA Curator', 'NASA Official: Jayne Dutra', 'Last Updated: July 15, 2007', and '+ Contact NASA'. A red arrow points from the 'Background & training materials' box to the top right corner of the page.

<http://web.archive.org/web/20110707202212/http://nasataxonomy.jpl.nasa.gov/2.0/>

# Demonstration of guided navigation of test collections



The screenshot displays the NASA website's search interface. At the top left is the NASA logo and the text "NATIONAL AERONAUTICS AND SPACE ADMINISTRATION". To the right is a search bar with the word "rover" entered and a "Go" button. Below the header, it states "219958 items". The main content area is organized into several columns, each with a category heading and a list of links with associated counts. The "by Collection" column is circled in red, and an arrow points from a text box to it.

**Search**

**219958 items**

|                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                      |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>by Organization</b> <ul style="list-style-type: none"><li><a href="#">NASA Affiliated Institutions</a> 1378</li><li><a href="#">NASA Centers</a> 76545</li><li><a href="#">NASA Contractors</a> 10108</li><li><a href="#">NASA Enterprises</a> 815</li><li><a href="#">NASA Headquarters</a> 4042</li><li><a href="#">Other NASA Partners</a> 999</li></ul> | <b>by Subject</b> <ul style="list-style-type: none"><li><a href="#">Aeronautics</a> 26532</li><li><a href="#">Astronautics</a> 31758</li><li><a href="#">Chemistry and Materials</a> 17086</li><li><a href="#">Engineering</a> 39631</li><li><a href="#">Geosciences</a> 30770</li><li><a href="#">Mathematical and Computer Sciences</a> 13286</li><li><a href="#">Space Sciences</a> 22685</li><li><a href="#">4 more</a></li></ul>                      | <b>by Missions and Projects</b> <ul style="list-style-type: none"><li><a href="#">Aerospace Technology</a> 60</li><li><a href="#">Biological and Physical Research</a> 68</li><li><a href="#">Data</a> 140</li><li><a href="#">Earth Sciences</a> 1497</li><li><a href="#">Human Exploration and Development...</a> 10680</li><li><a href="#">Planetary Missions</a> 4819</li><li><a href="#">Space Sciences</a> 9467</li></ul> | <b>by Date</b> <ul style="list-style-type: none"><li><a href="#">1972</a> 8392</li><li><a href="#">1973</a> 8512</li><li><a href="#">1974</a> 7828</li><li><a href="#">1975</a> 7704</li><li><a href="#">1992</a> 8131</li><li><a href="#">1993</a> 8519</li><li><a href="#">1994</a> 7712</li><li><a href="#">74 more</a></li></ul> |
| <b>by Competencies</b> <ul style="list-style-type: none"><li><a href="#">Business</a> 386</li><li><a href="#">Engineering</a> 393</li><li><a href="#">Mission</a> 555</li><li><a href="#">Scientific</a> 410</li><li><a href="#">Technical</a> 218</li></ul>                                                                                                   | <b>by Information Type</b> <ul style="list-style-type: none"><li><a href="#">Catalogs and Databases</a> 32</li><li><a href="#">Designs and Specifications</a> 62</li><li><a href="#">Plans and Agendas</a> 158</li><li><a href="#">Results and Analyses</a> 260</li><li><a href="#">Reviews and Lessons Learned</a> 1819</li><li><a href="#">Status Reports</a> 119</li><li><a href="#">Technical Reports</a> 229</li><li><a href="#">6 more</a></li></ul> | <b>by Collection</b> <ul style="list-style-type: none"><li><a href="#">Lessons Learned</a> 1370</li><li><a href="#">NTRS</a> 213900</li><li><a href="#">SIRTF</a> 4054</li><li><a href="#">Webb</a> 634</li></ul>                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                      |

Multiple collections viewed through a unified interface

uilt using Siderean Seamark (website and software are no longer available)

# IASA Taxonomy best practices

## ❖ Faceted classification schema

- Modular in nature for easier maintenance.
- Easy to tag what is appropriate to a use case.
- Facets give flexibility and power to search.

## ❖ Polyhierarchy

- Concepts can appear more than once.
- Enables knowledge discovery from multiple viewpoints.
- Fits a user-centric organization.

## NASA Taxonomy best practices (2)

### ❖ Hierarchical granularity

- NASA taxonomy is broad in nature by design.
- Different levels of depth depending on attribute set and content.
- Integration points allow for mapping of local vocabulary terms back to larger semantic framework.

### ❖ Use of standards

- Incorporates existing federal and industry terminology standards like NASA AFS, NASA CMS, FEA BRM, NAICS, and IEEE LOM.
- Complies with metadata standards like Z39.19, ISO 2709 (now ISO 25964), and Dublin Core.
- Political buy in from CIOs.

## NASA Taxonomy best practices (3)

### ❖ Mapping aliases

- Librarian approach allows for a meaningful expression of relationships.
- Synonyms, acronyms, related terms, broader and narrower terms
  - Example: MER A=Spirit, MER B=Opportunity
  - Example: Section 366 = Section 372
  - Example: JIMO = Jupiter Icy Moons Orbiter = Prometheus 1
- Easily maintained in derived RDF files.
- A means of holding institutional memory.

## NASA Specific Metadata Fields

*NASA has data sets unique to its business*

| Definition                   | NASA Taxonomy Mapping | Values                                      |
|------------------------------|-----------------------|---------------------------------------------|
| NASA missions and projects   | nasa:missionsProjects | Taxonomy: Missions and Projects             |
| Business purpose             | nasa:businessPurpose  | Taxonomy: Business Purpose                  |
| Technical competencies       | nasa:competencies     | Taxonomy: Competencies                      |
| Standard industry categories | nasa:industries       | Taxonomy: Industries<br>From NAICS standard |

## Project outcomes: NASA Taxonomy benefits

### .. at the NASA Level

- ❖ Metadata specification for all NASA content publishers
- ❖ RDF schema published with agreed standards (to enable appropriate use and reuse)
- ❖ Enhancement of Agency Web publishing processes
- ❖ Integration with NASA public portal content management system for:
  - Reduced publishing cycles
  - Better quality of Web materials – coordinated themes
- ❖ Integration with NASA Search Engine, Web Site Registration System
- ❖ Application in many technical areas, including engineering and science disciplines (STEP and science data dictionaries)

## Project outcomes: NASA Taxonomy benefits

### ... at the Federal Level

- ❖ NASA taxonomy development in accordance with e-Gov Act of 2002
- ❖ Integration with FEA at the BRM & 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

# White Papers and Supporting Documentation

- ❖ NASA Taxonomy
  - <http://web.archive.org/web/20110707202212/http://nasataxonomy.jpl.nasa.gov/2.0/>
- ❖ White Paper: Taxonomy Development With NASA, Dutra and Busch, 2003
  - <http://www.ppc.com/assets/pdf/case-studies/kmworld2010-day20-taxonomy-nasa.pdf>
- ❖ White Paper: Implementing the NASA Taxonomy Through Service Oriented Architectures, Dutra and Xiao, 2/2004
  - <http://taxonomystrategies.com/wp-content/uploads/2016/02/implementing-whitepaper.pdf>



# Summary

The National Aeronautics and Space Administration (NASA) is composed of more than a dozen separate Centers that engage in a variety of technological activities with missions and projects that reflect their specialized work. Technical communities across the Agency utilize highly evolved engineering and scientific vocabularies that reflect the nature of their disciplines. These vocabularies are often arcane and not easily translated by personnel outside a particular community of practice. This situation has tended to fragment information produced by NASA personnel. The goal behind building and adopting an agency-wide NASA Taxonomy [1] was to develop a consistent framework for handling NASA's electronic content, and to make it easy for various audiences to find all the relevant information from all the NASA programs quickly.

This talk looks back at a Taxonomy Strategies project to work with NASA from to develop a taxonomy and metadata standard meant to act as a classification scheme encompassing all of NASA Web content, including internal as well as external materials. The taxonomy provided a framework for tagging NASA Web material so that they could be used and reused in different applications such as the NASA portal and its supporting systems, the NASA search engine, and the NASA Web Site Registration System. In 2008, the Taxonomy Strategies was presented with an award from the NASA CIO for "providing major support on the NASA Taxonomy Task with expertise and knowledge on topics of vocabularies and metadata."

[1] <https://www.loc.gov/item/lcwa00099446/>