



# Taxonomy Use Examples

GSFC Project Librarians' Meeting

October 28, 2004

# Who I am: Joseph A. Busch



Over 25 years in the business of organized information

- ❖ Founder & Principal, Taxonomy Strategies
- ❖ Director, Solutions Architecture, Interwoven
- ❖ VP, Infoware, Metacode Technologies (acquired by Interwoven, November 2000)
- ❖ Program Manager, Getty Foundation
- ❖ Manager, Pricewaterhouse

Metadata and taxonomies community leadership

- ❖ President, American Society for Information Science & Technology
- ❖ Director, Dublin Core Metadata Initiative
- ❖ Adviser, National Research Council Computer Science and Telecommunications Board
- ❖ Reviewer, National Science Foundation Division of Information and Intelligent Systems
- ❖ Founder, Networked Knowledge Organization Systems/Services

# Recent Projects in Metadata and Taxonomy

## Government

- ❖ Defense Intelligence Agency
- ❖ Federal Aviation Administration
- ❖ FirstGov
- ❖ Forest Service
- ❖ HeadStart
- ❖ **NASA**
- ❖ Small Business Administration
- ❖ Social Security Administration
- ❖ USDA Economic Research Service
- ❖ USDA OCIO e-Government Program

## Non-Profit

- ❖ Dublin Core Metadata Initiative
- ❖ IDEAlliance

## Commercial

- ❖ Blue Shield of California
  - ❖ Halliburton
  - ❖ HP
  - ❖ Motorola
  - ❖ PeopleSoft
  - ❖ Sprint
  - ❖ Time, Inc.
- 
- ❖ for Critical Mass – Fortune 50 retail
  - ❖ for Deloitte Consulting – top credit card issuer

## International

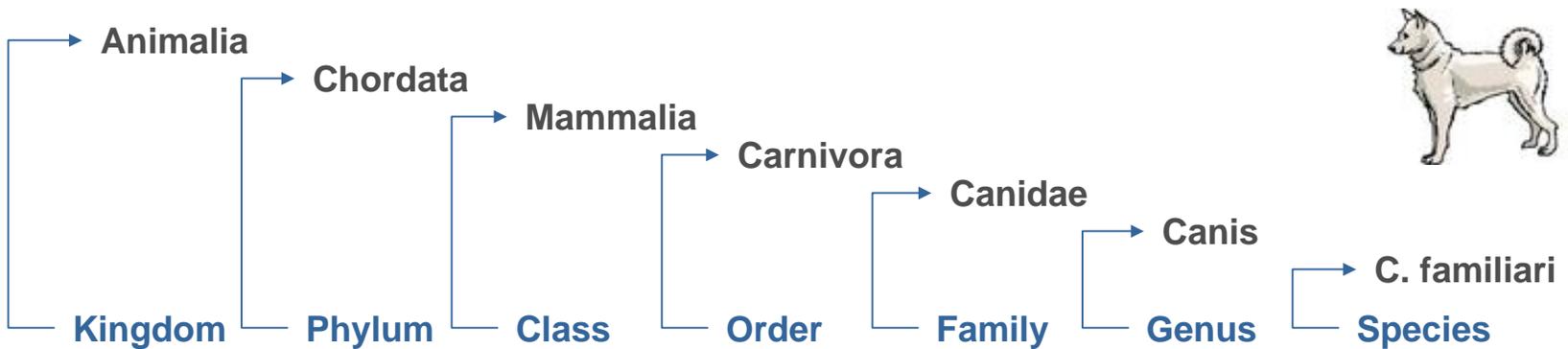
- ❖ European Standards Organisation
- ❖ Government of Singapore

# Metadata and Taxonomy Definitions

Terms	Definitions
Metadata	<p><u>Metadata is structured information to describe content.</u> Typical metadata fields are Title, Author, Subject, Publication Date, etc.</p>
Values	<p><u>Values for metadata fields may be free text</u> (e.g. Title), a specified data type such as a number or date, or come from a predefined list (e.g. predefined codes for Subjects).</p>
Controlled Vocabulary	<p><u>A managed set of terms that have been explicitly defined and agreed upon. All terms in a controlled vocabulary have an unambiguous, non-redundant definition.</u> Additions and deletions are “controlled”, meaning a process must be followed to change the list.</p>
Taxonomy	<p><u>The taxonomy is defined as a system for naming and organising things into groups that share similar characteristics.</u> It is a set of terms, organized into a structure. The terms might be the names of people, places, organizations, things, and concepts. The organization may be hierarchical and/or a set of mutually exclusive categories called facets.</p>
Facet	<p><u>Facets enable the classification of content from multiple dimensions.</u> It is a discrete branch of a taxonomy, with a separately maintained controlled vocabulary. Facet values are given in separate metadata fields.</p>

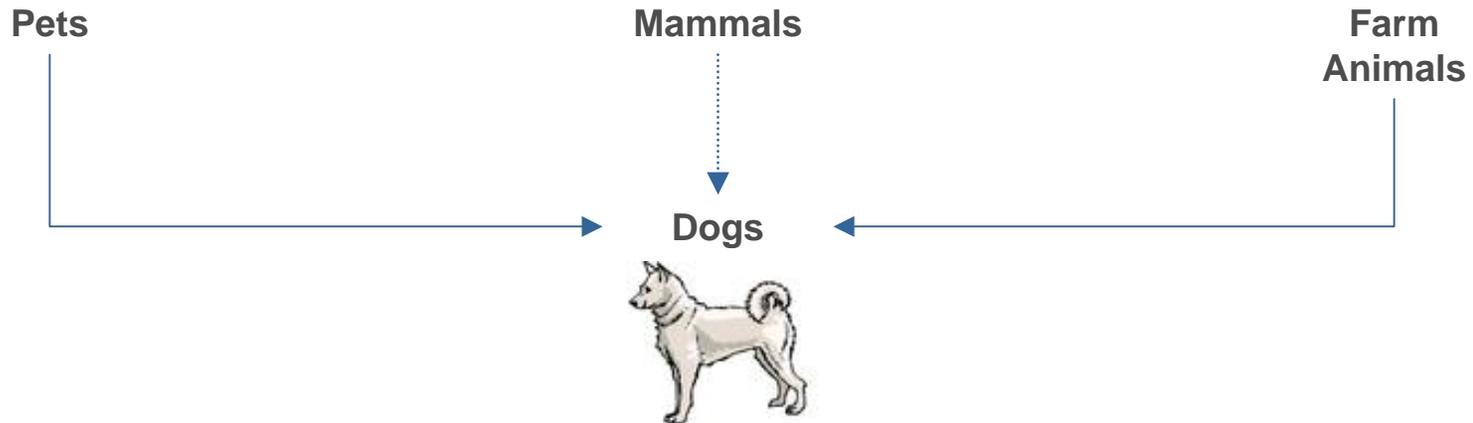
# Agenda

- ❖ Introductions
- ❖ What is Taxonomy?
- ❖ What is Dublin Core?
- ❖ What is the NASA Taxonomy?
- ❖ Using the NASA Taxonomy
- ❖ Case Study: JPL Unified Search for Project Information



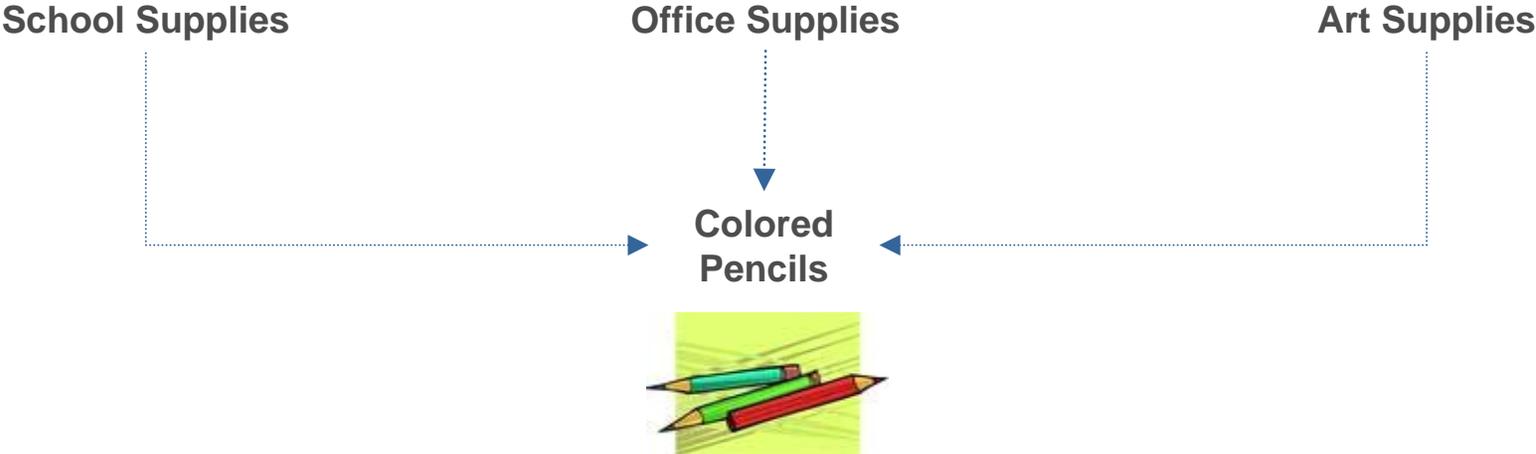
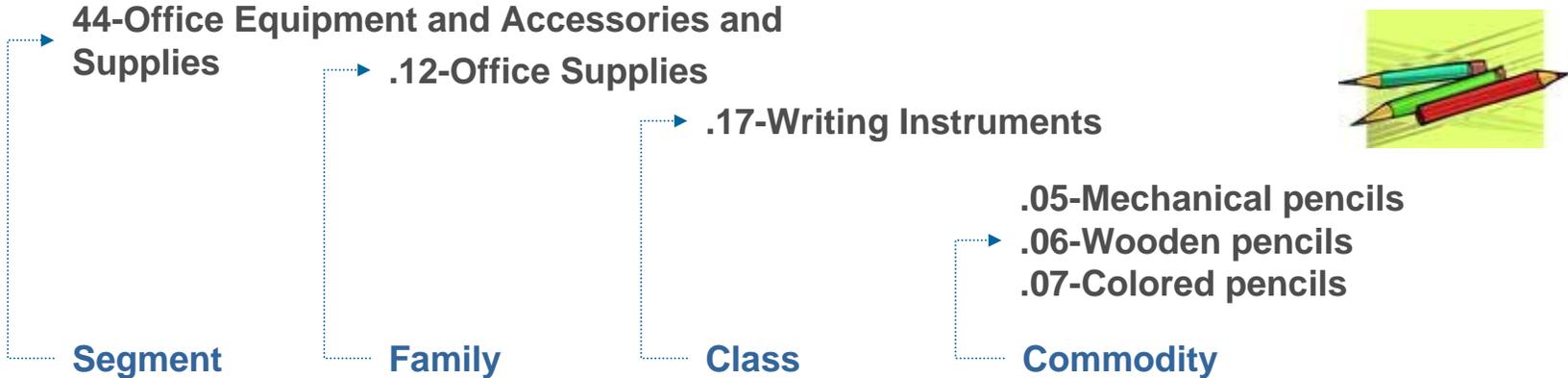
Linnaeus ...

**Biological taxonomy places an organism in one and only one place.**

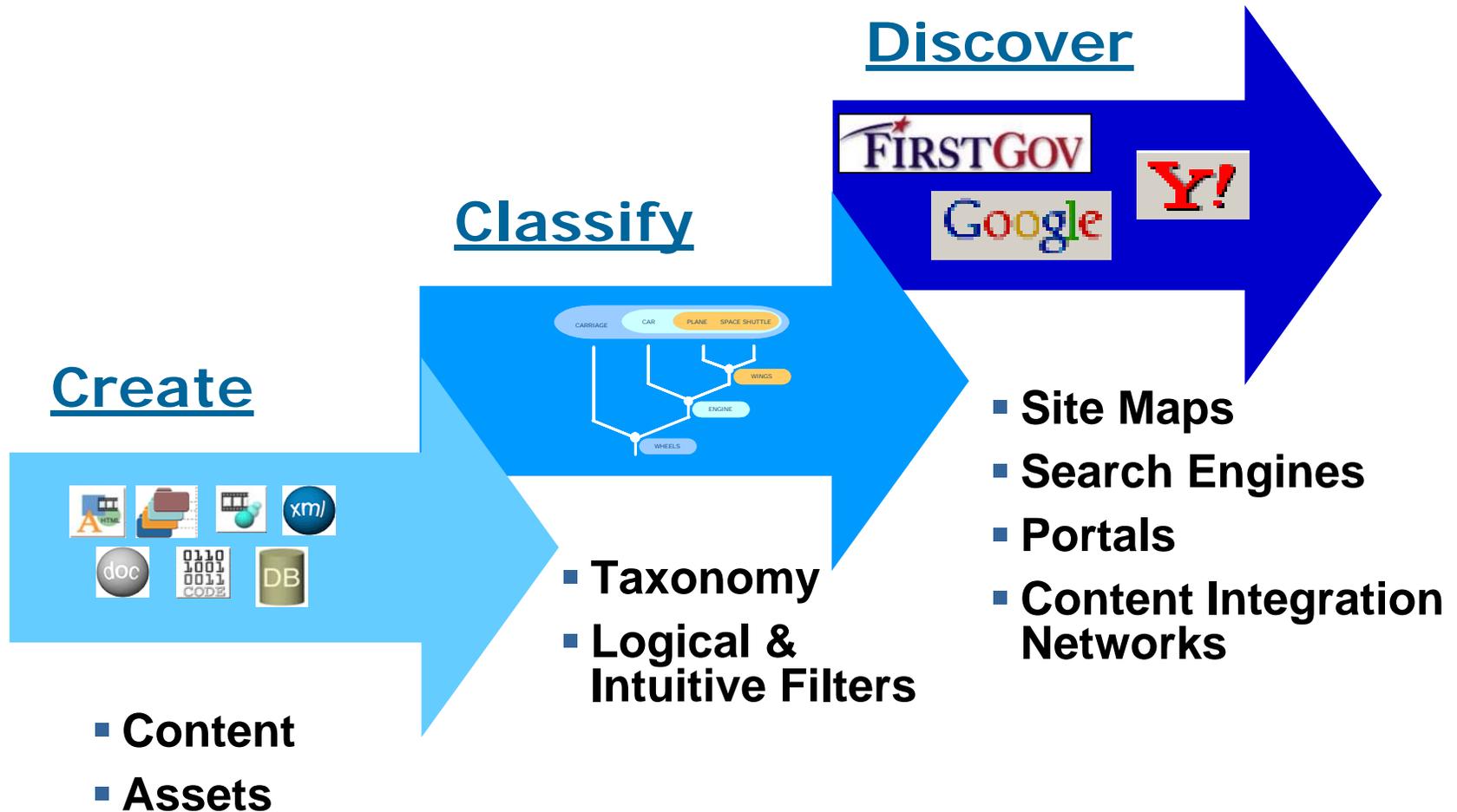


**But most of the time things belong to more than one category.**

# UNSPSC ...

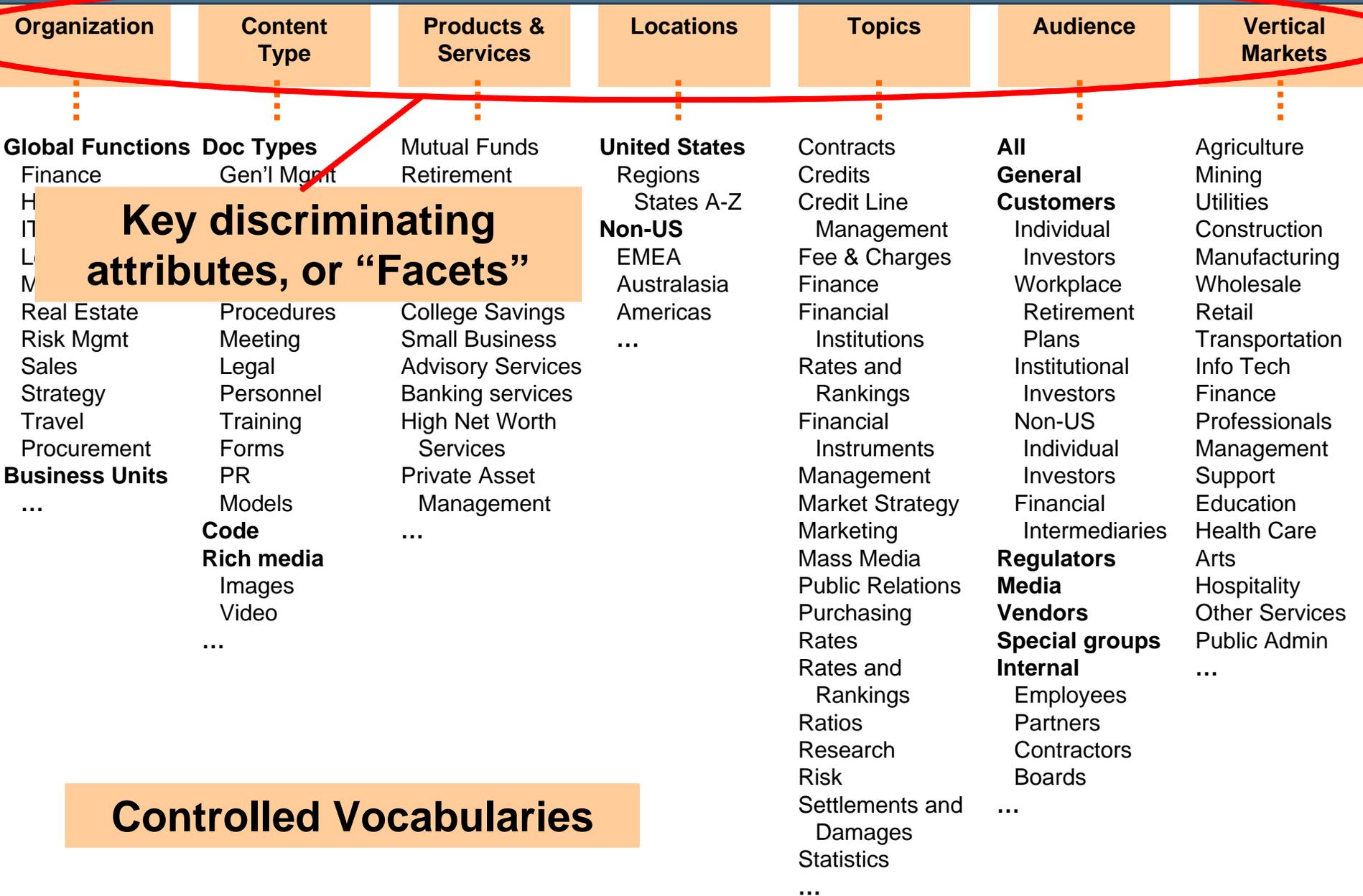


# What is the purpose of a Taxonomy? To ...



... find the right information at the right time to solve the problem at hand

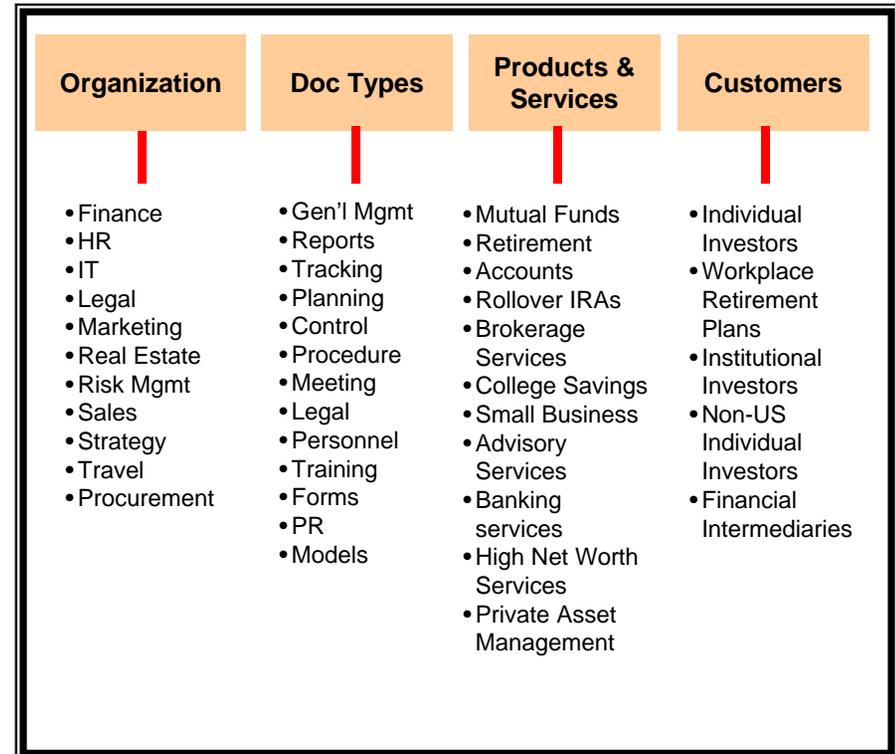
# Taxonomy



# The power of taxonomy facets

❖ 4 independent categories of 10 nodes each have the same discriminatory power as one hierarchy of 10,000 nodes ( $10^4$ )

- Easier to maintain
- Can be easier to navigate



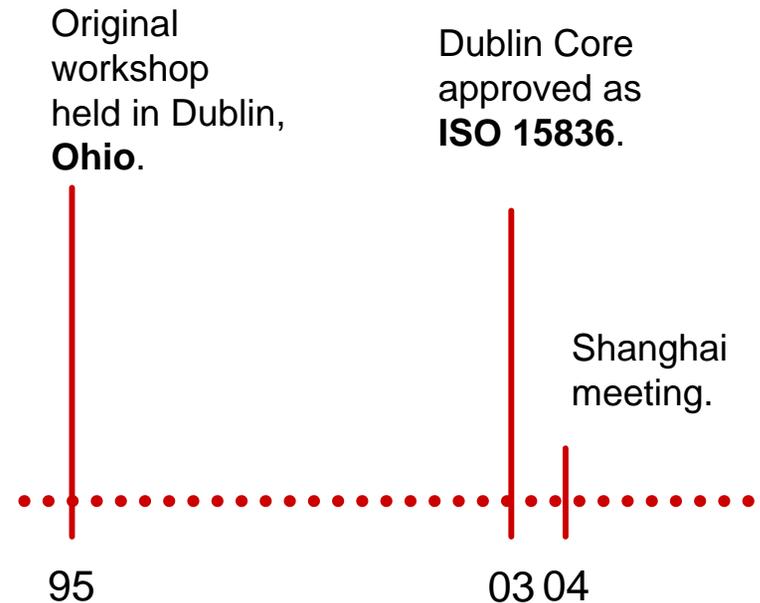
# Agenda

- ❖ Introductions
- ❖ What is Taxonomy?
- ❖ What is Dublin Core?
- ❖ What is the NASA Taxonomy?
- ❖ Using the NASA Taxonomy
- ❖ Case Study: JPL Unified Search for Project Information

# Dublin Core Metadata

- Dublin Core (DC) is the *Metadata Standard* for describing Internet resources so they are easy to find.
- DC is being used as the *starting point* for many metadata specifications.
- DC is a set of 15 basic elements. All are optional and repeatable.

Identifier	Date
Title	Source
Creator	Relation
Contributor	Rights
Publisher	Format
Subject	Type
Description	Language
Coverage	



For more information:

<http://www.dublincore.org>



# A Small Metadata and Taxonomy Example

## Metadata Standard

Field	Data Type / Source
Identifier	URL
Creator	string
Title	string
Date	date
Subject	Controlled Vocabulary or Taxonomy

## Example Taxonomy



The eCitizen.gov.sg website organized by topic into categories

### By Subject

#### Business

Intelligent Property, Business Assistance, Taxation

#### Housing & Property

Selling a property, Buying a Property, Housing Related Services

#### Defence & Security

Armed Forces, Border Protection, Rules & Regulations, Defence Policy

Culture, Recreation & Sport Clubs, Facilities, Fitness activities, Individual sports, Sporting events

Education, Learning & Employment Careers, Employment issues, Jobs, Education System, Training

Family, Community Development Activities, Culture, Community Development activities

#### Health & Environment

Health care, Hospitals, Health Groups, Environment studies

#### Transport & Travel

Traffic regulation, Vehicles and licenses, License types

#### Public Administration

National Day events, Institutes, Government Directories

## Metadata Standard Example

Field	Example Value
Identifier	www.iras.gov.sg/taxation/income_tax.html
Creator	MOF > IRAS
Title	Taxation in Singapore
Date	2002-01-11
Subject	Business > Taxation > Income Tax

# Agenda

- ❖ Introductions
- ❖ What is Taxonomy?
- ❖ What is Dublin Core?
- ❖ What is the NASA Taxonomy?
- ❖ Using the NASA Taxonomy
- ❖ Case Study: JPL Unified Search for Project Information

# NASA Taxonomy Project Progress

**PROJECT 3?**

## Follow-on Work

- Integrate with applications
- Long term maintenance

**PROJECT 2**  
Fall 03 - Spring 04

## Phase 4

- Dublin Core mapping
- Metadata specification
- XML schema development

## Phase 3

- Test & validate Taxonomy

## Phase 2

- Build community of practice
- Agree on comprehensive branches & taxonomy detail

## Phase 1

- Audit Content practices
- Identify & survey stakeholders

**PROJECT 1**  
Fall 2002

# NASA Taxonomy Web Site and Resource

The image shows a screenshot of the NASA Taxonomy website with several callout boxes pointing to specific elements:

- Link to Metadata Specification**: Points to the "+ NASA METADATA" menu item.
- Link to XML DTDs and Schema**: Points to the "+ NASA XML PROJECT" menu item.
- Background and training materials**: Points to the "Background and training materials" section.
- Links to Controlled Vocabularies**: Points to the "Subject Categories" link in the "NASA Taxonomy - Top Level Facets" list.

The website header includes the NASA logo and "NATIONAL AERONAUTICS AND SPACE ADMINISTRATION". A navigation bar contains: - TAXONOMY TOP LEVEL FACETS, + FAQs, + NASA METADATA, + NASA TAXONOMY XML, + NASA XML PROJECT. A sidebar on the right has: + Visit NASA.gov, + Contact the NASA Curator.

**NASA Taxonomy - Top Level Facets**

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

**Background and training materials**

Tips on using the NASA Taxonomy

**What is the NASA taxonomy?**

The NASA taxonomy is a controlled vocabulary that is designed to populate the [NASA metadata core specification](#).

It is also a means of tagging NASA content so that it can be used and reused in many different contexts.

Footer: FIRST GOV (Your First Click to the U.S. Government), + Freedom of Information Act, + The President's Management Agenda, + FY2002 Agency Performance Report. NASA logo, Contact the NASA Curator, NASA Official: Jayne Dutra, Last Updated: May 25, 2004.

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

# Jupiter's Ring System

<http://ringmaster.arc.nasa.gov/jupiter/jupiter.html#index>

## Reference Resource



Attribute	Values
<b>Information</b>	Web Sites; Animations; Images; Reference Sources
<b>Audiences</b>	Educators; Students
<b>Organizations</b>	Ames Research Center
<b>Missions and Projects</b>	Voyager; Galileo; Cassini; Hubble Space Telescope
<b>Industries</b>	N/A
<b>Locations</b>	Jupiter
<b>Functions</b>	Scientific and Technical Information
<b>Disciplines</b>	Planetary and Lunar Science
<b>Chronology</b>	1979-1999

# 2001 Mars Odyssey Data Archives

<http://wufs.wustl.edu/missions/odyssey/#Odyssey%20Data%20Sets>

## Data Archive

Attribute	Values
<b>Information</b>	Data Files; Web Sites
<b>Audiences</b>	Researchers; Scientists
<b>Organizations</b>	Jet Propulsion Laboratory
<b>Missions and Projects</b>	Mars Odyssey
<b>Industries</b>	N/A
<b>Locations</b>	Mars
<b>Functions</b>	Scientific and Technical Information
<b>Disciplines</b>	Planetary and Lunar Science
<b>Chronology</b>	2002-present



# Clementine – DSPSE

<http://www.cmf.nrl.navy.mil/clementine/>

## Web Site



Clementine - DSPSE



Clementine - DSPSE

Attribute	Values
<b>Information</b>	Web Sites; Data Files; Images
<b>Audiences</b>	Researchers; Scientists; Educators; Students
<b>Organizations</b>	Naval Research Laboratory
<b>Missions and Projects</b>	Clementine
<b>Industries</b>	N/A
<b>Locations</b>	The Moon
<b>Functions</b>	Scientific and Technical Information
<b>Disciplines</b>	Planetary and Lunar Science
<b>Chronology</b>	1994

# Using the NASA Taxonomy

- ❖ Do You Have to Use All Fields in Every Case?
- ❖ No, use what is appropriate to the case at hand.

# Why Do We Need So Many Metadata Fields?

The NASA Taxonomy is designed to be used in many different scenarios.

- ❖ 5\_Use Case Scenarios

# Use Case Scenario 1

## **Publishing to the NASA public portal.**

Could use these NASA Taxonomy elements:

- ❖ Audience
- ❖ Content Type
- ❖ Coverage (might be regional)
- ❖ Mission/Project
- ❖ Subject

# Use Case Scenario 2

## **Publishing to a NASA engineering portal.**

Could use these NASA Taxonomy elements:

- ❖ Audience
- ❖ Competency
- ❖ Content Type
- ❖ Mission/Project
- ❖ Subject
- ❖ Instrument

## Use Case Scenario 3

**Integrate information across multiple Centers for management reporting.**

Could use these NASA Taxonomy elements:

- ❖ Competency
- ❖ Content Type
- ❖ Mission/Project
- ❖ Subject
- ❖ Business Purpose

# Use Case Scenario 4

## Records Retention and Archiving.

Could use these NASA Taxonomy elements

- ❖ Building our knowledge base
- ❖ Assist in browse and navigation through large collections of disparate information objects

## Use Case Scenario 5

### **Query multiple repositories with a single, unified interface**

Could use these NASA Taxonomy elements:

- ❖ Spacecraft anomalies may require research across Problem Failure Reporting systems, PDMS systems, risk management databases, and many others
- ❖ Must be confident that all of the relevant material has been found

# Agenda

- ❖ Introductions
- ❖ What is Taxonomy?
- ❖ What is Dublin Core?
- ❖ What is the NASA Taxonomy?
- ❖ Using the NASA Taxonomy
- ❖ Case Study: JPL Unified Search for Project Information

# UNIFIED SEARCH for PROJECT INFORMATION

Search

2,232 items

## by Date

[1995](#) 202  
[1996](#) 322  
[1997](#) 245  
[1998](#) 148  
[1999](#) 187  
[2000](#) 126  
[2001](#) 152  
[2002](#) 208  
[2003](#) 354  
[2004](#) 129  
[3 more](#)

## by System

[Cassini Operations System \(COS\)](#) 848  
[Project](#) 4  
[Science](#) 9

## by Content Type

[Correspondence](#) 199  
[Designs and Specifications](#) 1355  
[Manuals](#) 41  
[Planning Documents](#) 128  
[Policies and Procedures](#) 356  
[Records](#) 4  
[Reports](#) 81

## by Project Level

[Ground System Level 3](#) 20  
[Ground System Level 4](#) 766  
[Ground System Level 5](#) 54  
[Ground System Level 6](#) 27  
[N/A](#) 13

## by Missions and Projects

[Cassini](#) 756  
[M98L](#) 61  
[Pathfinder](#) 499

## by Responsible Team/Group

[Data and Computing Services \(DCS\)](#) 34  
[DSN Services \(DSN\)](#) 3  
[Instrument Operations \(IO\)](#) 277  
[Mission Support and Services Offi...](#) 35  
[Real Time Operations \(RTO\)](#) 36  
[Science Planning \(SP\)](#) 3  
[Simulation and Verification Servi...](#) 8  
[Spacecraft Operations \(SCO\)](#)

## by Subsystem

[AIM](#) 464  
[Cassini](#) 2  
[DEFC](#) 20  
[DSCI](#) 3  
[EFC](#) 40  
[MARD](#) 8  
[MR](#) 2  
[PPS](#) 4  
[TELE](#) 6  
[UKN](#) 10  
[8 more](#)

## by Collection

[Cassini Electronic Library](#) 916  
[Problem Reporting](#) 1316

Can add as many repositories as needed for unified project search

# Results of 2003 JPL Information Repositories Study

- ❖ Fragmented and non-interoperable repositories
- ❖ Inefficient and broken processes and applications
- ❖ Parallel and redundant efforts both in building information systems and managing data
- ❖ Limited tools and services that cut across program and line organizations

<b>Data Repositories Identified</b>	
<b>Engineering</b>	<b>86</b>
<b>Science</b>	<b>8</b>
<b>Business/Admin.</b>	<b>28</b>
<b>Infrastructure</b>	<b>28</b>
<b>Outreach</b>	<b>1</b>
<b>Total</b>	<b>157</b>

# Semantic Frameworks Enhance Enterprise Data Architecture

- ❖ JPL data dictionaries are too narrow to interoperate
- ❖ EAs seeking “data harmonization”
- ❖ Semantic frameworks allow for mappings of data elements to larger vocabularies
  - Semantic relationships require more than simple controlled vocabularies
  - RDF statements allow specification of relationships for rules based inferencing

# Some Possible Semantic Engineering Solution Enablers

- ❖ An integrated semantic architecture that mirrors and extends the enterprise architecture
- ❖ JPL Taxonomy: controlled vocabularies for JPL engineering communities
  - By discipline, product, and process, etc.
- ❖ Centrally managed authority files for significant JPL asset attributes (ie, project names, etc)
- ❖ JPL Technical Thesaurus – equivalencies documented in RDF files
- ❖ Use semantic tools to present a unified navigation and search capability through JPL repositories

# Unified Project Search

## Integrating JPL Engineering Repositories

*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
  - ECR, PFRS, Docushare, Team Center, et. al.
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: **PFR** 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

# UNIFIED SEARCH for PROJECT INFORMATION

[How To Use This Tool](#)  
[Acronyms List](#)

Search

Resources: Cassini Electronic Library, Problem Failure Reporting System

2,232 items

## PFR Anomaly Type

### by Date

[1995](#) 202  
[1996](#) 322  
[1997](#) 245  
[1998](#) 148  
[1999](#) 187  
[2000](#) 126  
[2001](#) 152  
[2002](#) 208  
[2003](#) 354  
[2004](#) 129  
[3 more](#)

### by Subsystem

[AIM](#) 464  
[Cassini](#) 2  
[DEFC](#) 20  
[DSCI](#) 3  
[EFC](#) 40  
[MARD](#) 8  
[MR](#) 2  
[PPS](#) 4  
[TELE](#) 6  
[UKN](#) 10

### by Anomaly Type

[Incident Surprise Anomaly \(ISA\)](#) 692  
[LOG](#) 1  
[Problem Failure Report \(PFR\)](#) 567  
[Problem Log \(LOG\)](#) 3

### by System

[Cassini Operations System \(COS\)](#) 848  
[Project](#) 4  
[Science](#) 9

### by Content Type

[Correspondence](#) 199  
[Design Description](#) 34  
[N/A](#) 11  
[Operations](#) 4  
[Plans](#) 56  
[Procedures](#) 356  
[Reports](#) 80  
[Requirements](#) 57  
[Test Data](#) 72  
[User Guides](#) 37  
[4 more](#)

### by Project Level

[Ground System Level 3](#) 20  
[Ground System Level 4](#) 766  
[Ground System Level 5](#) 54  
[Ground System Level 6](#) 27  
[N/A](#) 13  
[TBD](#) 9

### by Missions and Projects

[Cassini](#) 756  
[M98L](#) 61  
[Pathfinder](#) 499

### by Responsible Team/Group

[Data and Computing Services \(DCS\)](#) 34  
[DSN Services \(DSN\)](#) 3  
[Instrument Operations \(IO\)](#) 277  
[Mission Support and Services Offi...](#) 35  
[Real Time Operations \(RTO\)](#) 36  
[Science Planning \(SP\)](#) 3  
[Simulation and Verification Servi](#) 8

# UNIFIED SEARCH for PROJECT INFORMATION

[How To Use This Tool](#)  
[Acronyms List](#)

## CEL and NASA Taxonomy Content Type

Resources: Cassini Electronic Library, Problem Failure Reporting System

2,232 items

### by Date

[1995](#) 202  
[1996](#) 322  
[1997](#) 245  
[1998](#) 148  
[1999](#) 187  
[2000](#) 126  
[2001](#) 152  
[2002](#) 208  
[2003](#) 354  
[2004](#) 129  
[3 more](#)

### by Subsystem

[AIM](#) 464  
[Cassini](#) 2  
[DEFC](#) 20  
[DSCI](#) 3  
[EFC](#) 40  
[MARD](#) 8  
[MR](#) 2  
[PPS](#) 4  
[TELE](#) 6  
[UKN](#) 10

### by Anomaly Type

[Incident Surprise Anomaly \(ISA\)](#)  
692  
[LOG](#) 1  
[Problem Failure Report \(PFR\)](#)  
567  
[Problem Log \(LOG\)](#) 3

### by System

[Cassini Operations System \(COS\)](#) 848  
[Project](#) 4  
[Science](#) 9

### by Content Type

[Correspondence](#) 199  
[Design Description](#) 34  
[N/A](#) 11  
[Operations](#) 4  
[Plans](#) 56  
[Procedures](#) 356  
[Reports](#) 80  
[Requirements](#) 57  
[Test Data](#) 72  
[User Guides](#) 37  
[4 more](#)

### by Project Level

[Ground System Level 3](#) 20  
[Ground System Level 4](#) 766  
[Ground System Level 5](#) 54  
[Ground System Level 6](#) 27  
[N/A](#) 13  
[TBD](#) 9

### by Missions and Projects

[Cassini](#) 756  
[M98L](#) 61  
[Pathfinder](#) 499

### by Responsible Team/Group

[Data and Computing Services \(DCS\)](#) 34  
[DSN Services \(DSN\)](#) 3  
[Instrument Operations \(IO\)](#) 277  
[Mission Support and Services Offi...](#) 35  
[Real Time Operations \(RTO\)](#) 36  
[Science Planning \(SP\)](#) 3  
[Simulation and Verification Servi](#) 8

# NASA Taxonomy Transitioned to a JPL Taxonomy





## UNIFIED SEARCH for PROJECT INFORMATION

# PFR Content Types merged with NASA Taxonomy to create JPL Taxonomy

2,232 items

### by Date

[1995](#) 202  
[1996](#) 322  
[1997](#) 245  
[1998](#) 148  
[1999](#) 187  
[2000](#) 126  
[2001](#) 152  
[2002](#) 208  
[2003](#) 354  
[2004](#) 129  
[3 more](#)

### by System

[Cassini Operations System \(COS\)](#) 848  
[Project](#) 4  
[Science](#) 9

### by Content Type

[Correspondence](#) 199  
[Designs and Specifications](#) 1355  
[Manuals](#) 41  
[Planning Documents](#) 128  
[Policies and Procedures](#) 356  
[Records](#) 4  
[Reports](#) 81

### by Project Level

[Ground System Level 3](#) 20  
[Ground System Level 4](#) 766  
[Ground System Level 5](#) 54  
[Ground System Level 6](#) 27  
[N/A](#) 13  
[TBD](#) 9

### by Missions and Projects

[Cassini](#) 756  
[M98L](#) 61  
[Pathfinder](#) 499

### by Responsible Team/Group

[Data and Computing Services \(DCS\)](#) 34  
[DSN Services \(DSN\)](#) 3  
[Instrument Operations \(IO\)](#) 277  
[Mission Support and Services Offi...](#) 35  
[Real Time Operations \(RTO\)](#) 36  
[Science Planning \(SP\)](#) 3  
[Simulation and Verification Servi...](#) 8  
[Spacecraft Operations \(SCO\)](#)

### by Subsystem

[AIM](#) 464  
[Cassini](#) 2  
[DEFC](#) 20  
[DSCI](#) 3  
[EFC](#) 40  
[MARD](#) 8  
[MR](#) 2  
[PPS](#) 4  
[TELE](#) 6  
[UKN](#) 10  
[8 more](#)

### by Collection

[Cassini Electronic Library](#) 916  
[Problem Reporting](#) 1316

# Cassini Data Rationalization

Collections: PFR System and 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**

# 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
- ❖ DDCR, Flight Project Practices

***There are many un-mined sources for semantic processing!***

# Questions and Discussion

## Contact Information:

Jayne Dutra, JPL

[Jayne.E.Dutra@jpl.nasa.gov](mailto:Jayne.E.Dutra@jpl.nasa.gov)

(818) 354-6948

Joseph Busch, Taxonomy Strategies

[jbusch@taxonomystrategies.com](mailto:jbusch@taxonomystrategies.com)

(415) 377-7912