Developing and Applying Controlled Vocabularies in Large Organizations for Increased Business Value

Jayne Dutra
JPL Information Architecture
NASA Taxonomy Manager
Semantic Technology Conference
March 2005

Publishing Cycle in the Real Time Organization

Create

- Content
- Assets

Classify

- Logical & Intuitive Filters
- Taxonomy

Discover

- Search Engine
- NASA Portals
- Content Integration Networks
- Enhance Enterprise Architecture

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

Project Benefits: Enable Knowledge Discovery

- Make it easy for users to find relevant information from NASA programs quickly
  - Provide easy access for NASA Web resources
  - Information integration for unified queries and management reporting
  - Enable semantic search across repositories
  - Enable the ability to move content through the enterprise to where it is needed most
  - Facilitate Records Retention and Management
- Comply with E-Government Act of 2002

NASA Taxonomy Best Practices

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

NASA Taxonomy Best Practices

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

NASA Taxonomy Best Practices

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, and Dublin Core
- Political buy in from domain experts and Communities of Interest

NASA Taxonomy Best Practices

Mapping Aliases

- Librarian approach allows for a meaningful expression of relationships
- Synonyms, acronyms, related terms, broader and narrower terms
- Easily maintained in derived RDF files

- *Example: MER A=Spirit, MER B=Opportunity*
- *Example: JIMO = Jupiter Icy Moons Orbiter = Prometheus 1*

A means of holding our institutional memory

Expressing Strategic Value With Semantic Frameworks

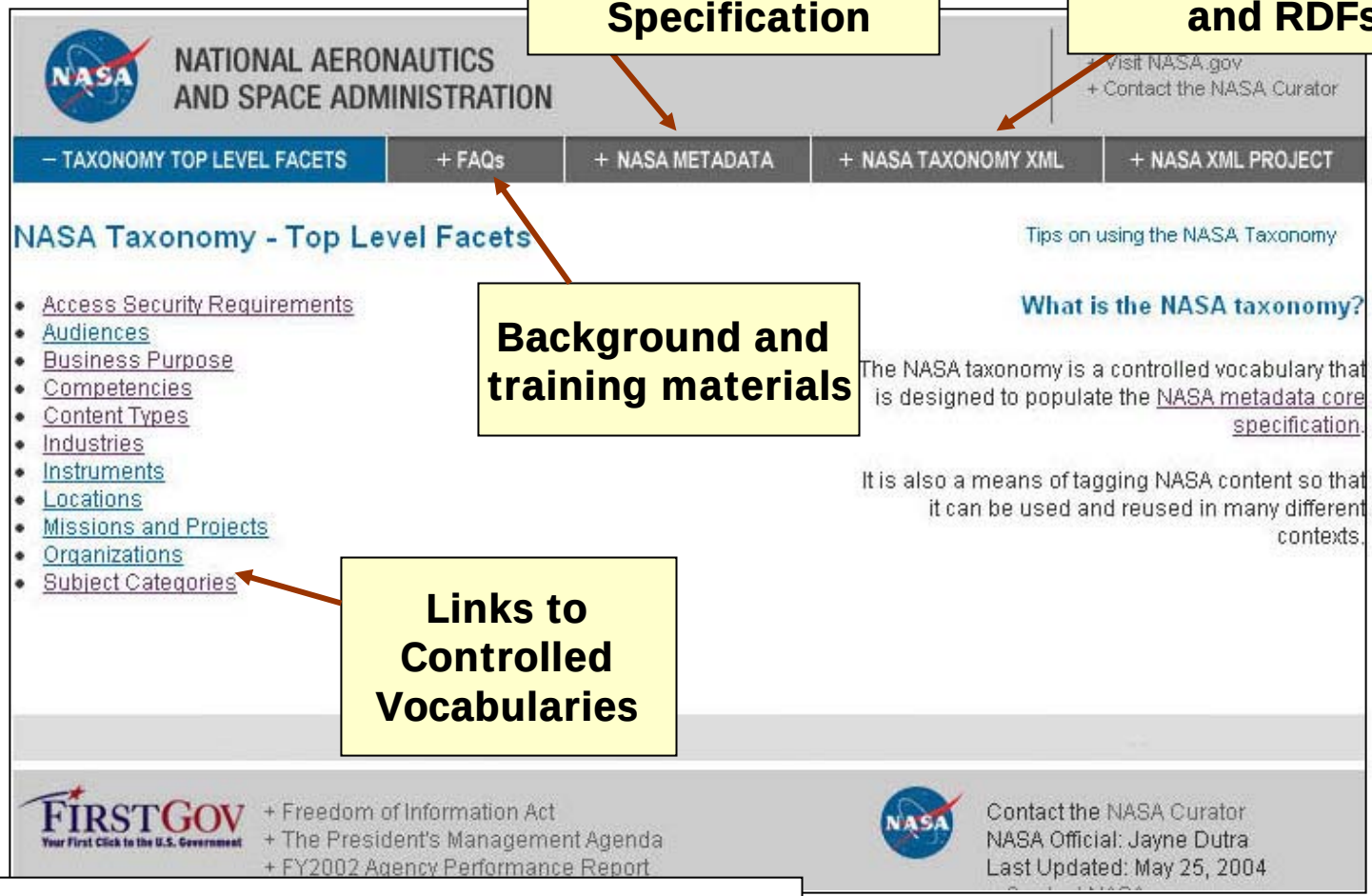
Facets	Strategic Value
Access Requirements	Sensitivity and access control
Audiences	Who is the content intended for
Business Purpose	Why the content was created
Competencies	Relevant field or discipline
Content Types	The genre of the content
Industries	External partners & businesses
Instruments	Flight 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 Specific Metadata Fields

NASA has data sets unique to its business

Definition	NASA Taxonomy Mapping	Value Sets
NASA missions and projects	nasa:missionsProjects	Missions and Projects
Business purpose	nasa:businessPurpose	Business Purpose
Technical competencies	nasa:competencies	Competencies
Standard industry categories	nasa:industries	Industries *From NAICS standard*

NASA Taxonomy Website and Resource



Link to Metadata Specification

Link to XML DTDs and RDFs

Background and training materials

Links to Controlled Vocabularies

http://nasataxonomy.jpl.nasa.gov

Project Outcomes: NASA Taxonomy Benefits

- Metadata specification for all NASA content publishers
- RDF schema published with agreed standards (to enable appropriate use and reuse)
- Semantic enhancement of search engines
- Integration with NASA portal content management system
- Integration with Web Site Registration System
- Application in many technical areas, including engineering and science disciplines (STEP and science data dictionaries)

A Case Study

Applying Controlled Vocabularies to Develop Semantic Search for JPL Flight Projects

JPL Projects Today

Parts Catalogues

Engineering Repositories

Electronic Libraries

Where do I find it?

Problem Reporting System

E-Mail Archives

Financial Databases

Semantic Frameworks Enhance Enterprise Information Architecture

- 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
- 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 and richer use

Semantic Engineering Solution Enablers

- An integrated **semantic architecture** that mirrors the enterprise architecture and is made up of several semantic components:
  - Common Metadata specification
    - Core Metadata Specification for JPL Project Documentation
  - Common language or controlled vocabularies
    - By discipline, product, and process, etc.
    - NASA Taxonomy, JPL Taxonomy
    - Ontologies of various knowledge domains
  - Standards and registries for consistency and discovery

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
   - Problem Failure Reporting System, Electronic Libraries, PDMS, Risk Management DB, 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: 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

UNIFIED SEARCH for PROJECT INFORMATION

How To Use This Tool
Acronyms List

Search [ ] Go

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 Anomaly Type
- Incident Suprise Anomaly (ISA) 692
- LOG 1
- Problem Failure Report (PFR) 567
- Problem Log (LOG) 3

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 Missions and Projects
- Cassini 756
- M98L 61
- Pathfinder 499

by Subsystem
- AIM 464
- Cassini 2
- DEFC 20
- DSCI 3
- EFC 40
- MARD 8
- MR 2

by System
- Cassini Oeprations System (COS) 848
- Project 4
- Science 9

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 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

Siderean Software tool Seamark used for faceted browse and navigation

UNIFIED SEARCH for PROJECT INFORMATION

How To Use This Tool
Acronyms List

Search Go

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 Anomaly Type
- Incident Suprise Anomaly (ISA) 692
- LOG 1
- Problem Failure Report (PFR) 567
- Problem Log (LOG) 3

by Content Type
- Correspondence 199
- Design Description 34
- N/A 11
- Operations 4
- Plans 56
- Procedures 356
- Reports 80
- Requirements 5
- Test Data 72
- User Guides 37
- *4 more*

by Missions and Projects
- Cassini 756
- M98L 61
- Pathfinder 499

CEL and NASA Taxonomy Content Type

by Subsystem
- AIM 464
- Cassini 2
- DEFC 20
- DSCI 3
- EFC 40
- MARD 8
- MR 2
- PPS 4
- TELE 6
- UKN 10

by System
- Cassini Oeprations System (COS) 848
- Project 4
- Science 9

by Project
- 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

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

Search Go

Resources: Cassini Electronic Library, Problem Failure Reporting System

PFR Anomaly Type

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 Anomaly Type
- Incident Suprise Anomaly (ISA) 692
- LOG 1
- Problem Failure Report (PFR) 567
- Problem Log (LOG) 3

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 Missions and Projects
- Cassini 756
- M98L 61
- Pathfinder 499

by Subsystem
- AIM 464
- Cassini 2
- DEFC 20
- DSCI 3
- EFC 40
- MARD 8
- MR 2
- PPS 4
- TELE 6
- UKN 10

by System
- Cassini Oeprations System (COS) 848
- Project 4
- Science 9

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 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

NASA Taxonomy

Content Types

- Designs and Specifications
- Quality Control

JPL Taxonomy

- Problem Failure Report
- Incident Surprise Anomaly
- Corrective Action Notice

Combined through RDF

Reconciling Schema

35	**Designs and Specifications**				
	Broader Term:	Content Types			
	Narrower Terms:	Configuration Controls			
		Notebooks			
		Quality Control			
				Problem Failure Report	2001
				Failure Report	2002
				Incident Surprise Anomaly	2003
				Problem Log	2004
				Corrective Action Notice	2005
				Discrepancy Report	2006
		Requirements			
	Term Number:		24		

NASA Taxonomy terms for engineering documents

JPL Taxonomy term extensions for engineering documents

UNIFIED SEARCH for PROJECT INFORMATION

Search DESCRIPTION Go

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 Content Type
- Correspondence 199
- Designs and Specifications 1355
- Manuals 41
- Planning Documents 128
- Policies and Procedures 356
- Records 4
- Reports 81

PFR Content Types merged with NASA Taxonomy to create JPL Taxonomy

- EFC 40
- MARD 8
- MR 2
- PPS 4
- TELE 6
- UKN 10
- *8 more*

by System
- Cassini Oeprations System (COS) 848
- Project 4
- Science 9

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 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
- cience Planning (SP) 3
- imulation and Verification ervi... 8
- pacecraft Operations (SCO)

by Collection
- Cassini Electronic Library 916
- Problem Reporting 1316

Mapping local vocabularies back to a common framework

UNIFIED SEARCH for PROJECT INFORMATION

Search DESCRIPTION Go

1,355 items matching
Content Type: Designs and Specifications ⊠

Refine

by Text Search
DESCRIPTION
Go

⊟ by Date
1994 70
1995 194
1996 291
1997 108
1998 100
2000 98
2001 75
2002 133
2003 122
2004 57
3 more

⊟ by Content Type
Quality Control 1263
Requirements 58

⊟ by Missions and Projects

1 to 10 of 1,355 items | next

Content Type: Requirements
System: Cassini Oeprations System (COS)
Project Level: Ground System Level 3
Responsible Team/Group: Simulation and Verification Services (SVS)
TITLE: 699-500-SVSHS/R - Ground System Requirements on the High Speed Spacecraft Simulator
Collection: Cassini Electronic Library

Date: 1997
Content Type: Requirements
System: Cassini Oeprations System (COS)

(SCO)
trol System/Propulsion FDD/FRD

(SCO)
Collection: Cassini Electronic Library

Date: 1996
Content Type: Requirements
System: Cassini Oeprations System (COS)
Project Level: Ground System Level 5
Responsible Team/Group: Spacecraft Operations (SCO)
TITLE: 699-FSOFDD-504 - System/Command Data System/Telecom Team (SCT)

Quality Control records as a sub-type of Designs and Specifications from the NASA Taxonomy

Integrating Center Taxonomies to Create a Common NASA Vocabulary

JPL Problem Failure Reporting Content Types displayed as a Sub Term of Quality Control

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

Next terms mapped to each other using an existing spacecraft bus ontology.
Next facet added was Instruments.

Final Results of Data Harmonization for Cassini

- 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

Summary

- Controlled vocabularies together with semantic technologies offer us many new ways to integrate information that have not been available to us before

- Information models are enhanced by using RDF which allows us to specify relationships between terms and also concepts (ontologies)

Thanks For Your Time!

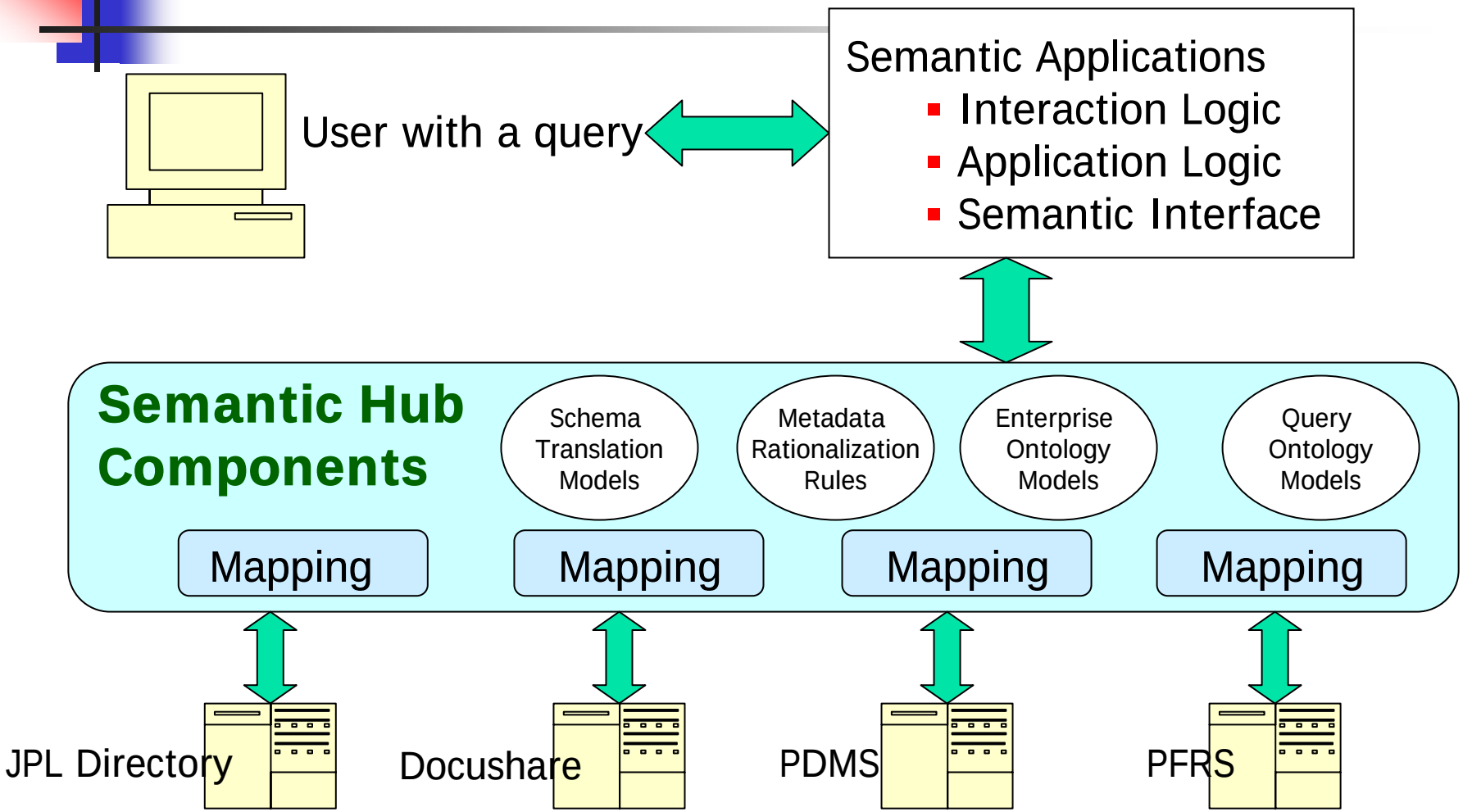
Jayne Dutra

Jet Propulsion Laboratory, NASA

Jayne.E.Dutra@jpl.nasa.gov

Back Up Slides

Notional Semantic Architecture



Some Business Rules for Schema Reconciliation Across Domains

Rule 1.

- Where a single, authoritative reference source defines the values that may be used
  - This is called “Harmonization”
  - Example: the names of JPL’s missions and projects
  - Held in a CVA (controlled vocabulary authority file) for use by developers
  - Could be available through a Metadata Registry

Some Business Rules for Reconciliation Across Domains

Rule 2.

- Where the characteristics of an attribute are consistent across sources
  - Example: *Author = Creator*
  - Example: *Responsible Team = WBS*

Or

- The notion of equivalencies
  - Example: *MER A=Spirit, MER B=Opportunity*
  - Example: *Section 366 = Section 372*

RDF Holds Equivalencies and Related Terms


```
- <pcv:Descriptor rdf:about="subj:63">
    <pcv:label>Cybernetics</pcv:label>
    <pcv:synonym>artificial intelligence</pcv:synonym>
    <pcv:synonym>control theory</pcv:synonym>
    <pcv:synonym>expert systems</pcv:synonym>
    <pcv:synonym>feedback theory</pcv:synonym>
    <pcv:synonym>robotics</pcv:synonym>
    <pcv:broaderTerm rdf:resource="subj:59" />
    <pcv:relatedTerm rdf:resource="subj:54" />
    <pcv:definition>Includes feedback and control theory, artificial intelligence, robotics and expert
      systems. For related information see also Man/System Technology and Life Support.</pcv:definition>
    <pcv:code>67</pcv:code>
  </pcv:Descriptor>
- <pcv:Descriptor rdf:about="subj:45">
    <pcv:label>Environment Pollution</pcv:label>
    <pcv:synonym>air pollution</pcv:synonym>
    <pcv:synonym>noise pollution</pcv:synonym>
    <pcv:synonym>pollution</pcv:synonym>
    <pcv:synonym>thermal pollution</pcv:synonym>
    <pcv:synonym>water pollution</pcv:synonym>
    <pcv:broaderTerm rdf:resource="subj:42" />
    <pcv:relatedTerm rdf:resource="subj:71" />
    <pcv:definition>Includes atmospheric, noise, thermal, and
    <pcv:code>8</pcv:code>
```


Defining Equivalent Terms in RDF

Acronyms are easily defined using Equivalent Terms or Synonyms in RDF

Some Business Rules for Reconciliation Across Domains

Rule 3.

- Where distinctions in metadata values do not add value for users
  - Example: Highly technical or very granular attributes important to a narrow set of domain users may not have much significance to a wider audience

Some Business Rules for Reconciliation Across Domains

Rule 4.

- Where rationalization makes search more efficient for users without compromising the integrity of the data
  - Example: Mapping Section 319 document types to the JPL or NASA Taxonomies

Distinctions between Project Content Types

	A	B	C
1	**Sec 319 Doc Type**	**Classified Subtype**	**NASA Taxonomy Mapping**
2	Agreement		Record:Agreement
3		MOU	Record:Agreement
4		MOA	Record:Agreement
5		ICD	Record:Agreement
6		Contract (Discreet)	Record:Agreement
7	AIDS		
8	Analysis		Reviews, Results and Analyses
9	Announcement of Opportunity		Announcement
10	Correspondence		Correspondence
11	Cost data (Discreet)		Financial Records
12	Design document		Designs and Specifications
13		Dictionary	Designs and Specifications
14		Matrix	

Agreements can be Memos of Agreement or Memos of Understanding