Developing Integrated Taxonomies for a Tiered Information Architecture

Jayne Dutra, Enterprise Information Architecture
Online Information 2006 Conference
November 30, 2006
Who Am I?

Core Capabilities:
- Search, web design, portal technology, web content management, knowledge management, information architecture, taxonomy development, metadata design, business analysis for IT applications

Past Credits:
- Inside JPL Portal (home page including enterprise search), Manager for NASA Taxonomy and Core Metadata Specification, Co-Chair of NASA Web Managers, JPL KM Process Owner, JPL Engineering Taxonomy and Metadata Core Spec Task Manager, team member: NASA Taxonomy for Problem Reporting Systems, faceted navigation pilot for flight projects

Current:
- Most of my work is for the CIO’s Office on EA tasks, with an emphasis on information architecture problems
Problem Statement: JPL Today

Parts Catalogues

Electronic Libraries

E-Mail Archives

Engineering Repositories

Problem Reporting

What did I call it?
Where did I put it?
How do I find it?
The Solution: An Integrated Enterprise Information Architecture

A Unified View of our information space

- Integrated with business processes and best practices
- Based on institutional policies and common architecture
- Independent of any specific repository or technology
- Able to deal with harmonization of information according to a larger point of view
- Relevant and useful to our customers

Expressed with a shareable metadata and its values (taxonomy)
Information Building Blocks

An integrated information architecture made up of several components:

- **Common Metadata Specification**
  - Core Metadata Specification for JPL Information Objects
- **Common language or controlled vocabularies**
  - By discipline, product, life cycle, process, etc.
  - NASA Taxonomy, JPL Taxonomy, Partner Taxonomies
- **Business Rules for data reconciliation**
  - You say “tomato”......
Nested Taxonomies

Term criteria at the Enterprise level

- Need for **broad** buckets of information that many different systems can map into
- Patterns tied to business process emerge over time
- Integrated with the NASA Taxonomy for interoperability
Tiered Taxonomy Work

- Enterprise Information Architecture is tiered
- Work already exists in the Project Engineering and Science Domains
- Work is starting up in the Business Domain
- Work is waiting in the Work Force Domain
Work Already Completed: Metadata Core Spec for the Project Domain

Sponsored by the JPL Office of the CIO

Goals

- Create Lab wide information standards for engineering products
- Design and document a Lab wide information content model
- Improve information access and retrieval, including cross repository search
- Integrated with NASA metadata and taxonomy specs

Deliverable

- JPL Core Metadata Specification for Project Documentation, v1.0
- Developed by a cross organizational team
- Delivered March, 2005
JPL Enterprise Project Taxonomy Effort Goals

- Tightly integrated with JPL Project Metadata Core Specification
  - Provides vocabulary values for the appropriate metadata fields
  - Designed to be relevant for a domain of users (communities of practice)

- Provides a classification scheme for identifying content
  - A means for tagging content so it can be used and reused in different contexts

- Derived from JPL standard gold sources
  - Vocabulary owners are Subject Matter Experts in their domains
  - Discoverable with SOAs or Semantic Web technologies
JPL Flight Project Engineering Taxonomy

Tips on using the JPL Taxonomy

What is the JPL taxonomy?

The JPL taxonomy is a controlled vocabulary that is designed to populate the JPL metadata core specification.

- Taxonomy values, equivalencies
- Metadata Core Specification
- RDF files for easy re-use
- FAQ files to assist developers
Adding Richness Over Time

- Metadata Core Specification was a top down effort
Adding Richness Over Time

- Metadata Core Specification was a top down effort
- For the Phoenix Faceted Navigation Pilot, we used a bottom up approach
Validation: Faceted Navigation Prototype for Engineers

Semantic Search/faceted navigation for Flight Projects

- JPL Engineering Domain
- Pilot funded by JPL Chief Engineer
- For flight teams on Phoenix and CloudSat missions
- Completed in 6 weeks
- Goal: Provide cross repository search from a single interface based on
  - Relationships of information objects
  - Life cycles – mission and content
  - Task analysis for specific roles
  - Engineering processes
Phoenix Semantic Search Pilot

- 15,000 items
- 5 repositories
- Tagged with JPL
- Taxonomy facets
- Unified search from one interface
- Faceted navigation for intuitive info discovery
- Sponsored by the JPL Chief Engineer
Some Further Use Cases: Faceted Navigation Using Relationships

- **Project Manager:**
  - “I’d like to check all documents needed to complete my Certification for Launch to see what state they’re in, no matter where they are.”

- **Cognizant Engineer:**
  - “I’d like to see all problem failure reports on a sub-system I designed and flew 5 years ago so I can incorporate the lessons learned into my current mission.”

- **Project Information Management Engineer:**
  - “I’d like to see the status of all Phase B documents that I need to prep for an upcoming gate review so I know we’re ready.”
Work In the Financial Arena

JPL Business Domain

- Term clarification for vocabulary used in financial reporting, work force planning, estimating and proposals

Sample Vocabulary Problem Term: Project
Oracle project number – financial costing system
IBF/NSM project number – NASA reference
PRL project number – JPL flight project name

How can we efficiently search for information about a “project”? 
### Describing Work for Forecasting and Strategic Planning

<table>
<thead>
<tr>
<th>Strategic Question</th>
<th>Metadata Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>What work are we doing?</em>&lt;br&gt; <em>What type of business is it?</em></td>
<td>- Work Type  &lt;br&gt;  - Business Type  &lt;br&gt;  - Work ID/ Project Name</td>
</tr>
<tr>
<td><em>How did we acquire it?</em></td>
<td>- Competition Type</td>
</tr>
<tr>
<td><em>Who is funding the work?</em>&lt;br&gt; <em>Who are we delivering it to?</em></td>
<td>- Role (inc customer, sponsor, partner, line roles and project roles)</td>
</tr>
<tr>
<td><em>Who is doing the work?</em>&lt;br&gt; <em>What capabilities does it take?</em></td>
<td>- Organization  &lt;br&gt;  - Competency, Capabilities  &lt;br&gt;  - Work description</td>
</tr>
<tr>
<td><em>What phase is it in?</em>&lt;br&gt; <em>Where is it taking place?</em></td>
<td>- Work State  &lt;br&gt;  - Work Maturation Level  &lt;br&gt;  - Implementation Mode</td>
</tr>
</tbody>
</table>
Work Already Completed: Metadata Working Group for Business Domain

Goals

- Create Lab wide information standards for business/work products
- Design a Lab wide business information content model
- Improve information access and retrieval, including cross repository search and reporting
- Identify “Gold Source” owners of data

Deliverable

- JPL Core Metadata Specification for Business, alpha version
- Developed by a cross organizational team
- Delivered May, 2006
- Next steps: validation and assigned owners
More work continuing with Integrated Business team for validation and expansion
Work Force Metadata

- Collect and document attributes about people

- Some use cases
  - Access management – who are you and what can you see?
  - Targeted content delivery – what content helps you get your work done?
  - Work force planning – what skill sets do you have that we can apply to work?
  - IT Services – How can we provision you with the proper IT services?
Using People Metadata for Targeted Content Delivery in Portals and Dashboards

- Make content available to delivery mechanisms using Service Oriented Architectures
- Data streams presented as services and available for consumption by workers in portals, dashboards and other devices

Source A

Source B

Source X, Source Y, Etc.

UDDI Registry
WSDL
SOAP, etc
Attributes That Describe People
- An Engineer
- Specialty is Electrical Engineering
- Works on propulsion systems
- Worked on projects X, Y, Z
- Currently working on A
- As a Cog E
- On propulsion subsystem
- Project is in Phase C
- Has published papers on propulsion systems

Corresponding Taxonomy Facet
- Competency
- Capability
- Topic or Subject Matter
- Past Assignments
- Current Assignment
- Role
- System/Subsystem
- Project Phase
- Topic or Subject Matter

We can codify and track certain attributes for re-use
Matching Attributes for People to Attributes for Content

Attributes About People
- Competency/Discipline
- Subject Matter
- Past Task Assignment - Role
- Current Task Assignment - Role
- Subsystem/WBS
- Task Phase
- Associations to objects as Author or Reader/Subscriber

Attributes About Info Objects
- Objects related to a Competency
- Interest in Subject Matter Areas
- Objects associated with Role
- Information on a Subsystem/WBS
- Objects associated with a project phase
- Information on project products
- Information on technologies

People Metadata

Project Metadata
Tiered Taxonomy Work

Enterprise Information Architecture is tiered

Enterprise Taxonomy

- Project Taxonomies (Products)
  - Engineering
  - Science
- Business Taxonomy (Work and $$)
- Work Force Taxonomy (People)
How An Enterprise Taxonomy is Born: The Cream Rises to the Top

Enterprise Taxonomy

- Organization
- Mission/Projects

Project Taxonomy
- Content Type
- Mission Phase
- Organization
- Document State
- Missions/Projects
- System/Subsystem

Business Taxonomy
- Work Type
- Funding Mechanism
- Business Type
- Organization
- Acquisition Type
- Missions/Projects

Work Force Taxonomy
- Organization
- Competency/Skills
- Capability
- Missions/Projects
- Role
Iterative process in a tiered model

- Direct connection between IA and use cases, business questions
- Metadata specification defined
- Taxonomy definition: Values expressed with controlled terms where appropriate
  - Domain specific taxonomies add business value
  - Transition points enable interoperability
  - Look for patterns of usage tied to business process
- Integrate across domains and upward to a meta level
- Implement and validate
- Long term stewardship and configuration management
Questions, Discussion

Thanks!

Jayne.E.Dutra@jpl.nasa.gov
Information and IT Services Delivery by Mission (and Content) Life Cycle Based on Role

Leverage Mission Life Cycle
- What roles are active at each stage
- What IT services are needed at each phase of a mission
- What capabilities are most in use at each phase of a mission

Example:
Proposal Phase → System Engineering Capability
Activities → Trade Studies, Model Building, Requirements Formulation
IT Needs → Document storage, simulation and visualization, traceability
IT Apps → Docushare, MBED, Cradle, CORE

Other Proposal Phase roles and activity types:
- Project Managers – do planning
- Principal Investigators – formulate science definition goals
- Project Resource Administrators – create budget and schedules

This is a very rich area for us to mine!
Notional IA Architecture – High Level

- **LDAP Metadata Catalogue**
  - Semantic Representations

- **SchemaLogic Server**
  - Taxonomy
  - Metadata Schemas

- **UDDI**
  - (phonebook of web services)

- **Search/Reporting/Portal Applications**

- **Metadata Catalogue**
  - Semantic Representations

- **Seamark Server**

- **Web Services Wrapper**

- **Repositories**
  - (Docushare, Rules!, PDMS, …)

- **Web Services Wrapper**

- **XML, RDF, XMLS, OWL**

- **RSS, XML, SOAP**

- **XRBR**

- **WSDL**
A view from the top . . .

IA work supports many different stakeholders

Priority
Improve our ability to work more efficiently

Goal
Improve ability to store, archive, retrieve project information

Processes
Enterprise Content Management
Product Lifecycle Management
Information Discovery and Retrieval

Capabilities
Document Storage
Web Content Management
Records Management
Work Flow

Product Data Management
Requirements Management
Risk Management

Cross Repository Retrieval
External Partners Data Exchange
Access Verification
Export Compliance

Technologies
Electronic Library - DocuShare
Document Repository - Teamcenter Community
Web Content - Rythmyx

PDMS - Teamcenter Enterprise
Requirements Repository - DOORS, Cradle, Core
Risk Management - ARM

Portals – Inside JPL, Teamcenter Community
Search Engine – Google
Problem Reporting - PFR/PRS
Manufacturing/Inventory - iPICS

Common Information Infrastructure
Security: Authentication
Metadata Standards

Unique Object Identifiers
Domain Taxonomies
Schema Registries
Unified Search for JPL Projects: Goal

Resulting in integrated responses like this:

- Project System Requirements v 2.0 (DOORS)
- Hardware Drawing - Parts v 1.0 – PDMS
- Hardware Drawing – Assemblies v 3.0 - PDMS
- ECR – PDMS
- ECR waivers – Team Center Community
- Problem Failure Report - PRS
- Associated Documentation - Docushare

Requires Semantic Web technology and Service Oriented Architectures To automate
Unified Search for JPL Projects: Goal

Users would like to be able to find related sets of data. For example: “I want to see all of the released documents for my project that are required for system certification for flight”

Combinations of metadata...
- Project = Project XYZ
- Document State = Preliminary or Official
- Content Class = Controlled
- Business Purpose = Product Data
- Content Type = Appropriate Values
- Document Status = Lifecycle values

Query can be applied across multiple repositories...
Taxonomy Packages Equivalencies for Vocabularies That Change Over Time

Acronyms and Synonyms can be easily defined using Equivalent Terms in RDF (a new language from the semantic web)

- MER A = Spirit, MER B = Opportunity
- MER A and MER B are types of Rovers
- Rovers fall under the subject heading of Robotics/Cybernetics

Users don’t have to have special contextual knowledge to find information items
Core Metadata Specification Online

Expressed with metadata and its values (taxonomy)

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Element Name and Namespace</th>
<th>Definition</th>
<th>Data Type or Source</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>dc:creator</td>
<td>Preparer or responsible individual</td>
<td>String</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>dc:date</td>
<td>Date the resource was 'published'</td>
<td>Cover Date, use format yyyy-mm-dd</td>
<td>For browsing by date, a simple publication date is all that is needed.</td>
</tr>
<tr>
<td>Class</td>
<td>jpl:class</td>
<td>Whether or not the content is under configuration management</td>
<td>Yes/No</td>
<td>If material is controlled, other conditional tags may apply</td>
</tr>
<tr>
<td>Content Type</td>
<td>dc:type</td>
<td>The nature or genre of the content of the resource</td>
<td>Values come from JPL Taxonomy: Content Types</td>
<td></td>
</tr>
<tr>
<td>File Name</td>
<td>jpl:file</td>
<td>Label given to a file that contains the content</td>
<td>String</td>
<td></td>
</tr>
</tbody>
</table>

This specification was developed with a focus on JPL flight project documentation. Excel Spreadsheet

http://jpltaxonomy/metadata.htm
Acronyms can be easily defined using Equivalent Terms or Synonyms in RDF

MER A= Spirit, MER B = Opportunity, MER A and MER B are types of Rovers and fall under the subject heading of Robotics/Cybernetics

Built into the back end and transparent to users
Answering important business questions for forecasting:

- What work are we doing? What type of business is it?
- How did we acquire it?
- What skill sets does it take to complete?
- Who is funding it? Who is doing it? Who are we delivering it to?
- How much does it cost? How many hours to completion?
- What phase is it in? Where are we doing it?

We want to improve how accurately and how quickly we can answer these questions for management.