The Semantic Web: Concepts, Deployment Options and Software Demonstration

Digital Asset Management Symposium
Los Angeles, November 15, 2005

Jayne Dutra
Jet Propulsion Laboratory,
California Institute of Technology

Christian Halaschek-Wiener
MINDSWAP Research Group
The University of Maryland
Who Are We?

Jayne Dutra
- Jet Propulsion Lab, NASA
- Manager, NASA Taxonomy
- Manager, JPL Taxonomy
- Information Architect
- Knowledge Management Process Owner
- *Fascinated by the Semantic Web!*

Christian Halaschek-Wiener
- MINDSWAP Research Group
  University of Maryland
- Graduate Research Assistant
- Multimedia Management on the Web
Session Agenda

- What is the Semantic Web?
- How does it work?
- What business value does it have?
- What can we do with it?
- Software Demo: Multimedia Annotation
From Tim Berners-Lee and the W3C

“The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications.”

http://www.w3.org/2001/sw/
So, What is the Semantic Web?

- Today’s Web is made for **people** to read and understand
- Tomorrow’s Web will be made for **computers** to read and understand
  - Systems will be able to perform transactions across applications without human help
  - Leverages the vast amount of data accessible on the Web for machine processing
  - Integration of data sets that are currently unlinked using the Web
How Does It Work?

- Focused on encoding metadata about Web resources into Web pages
  - Good to start with a basic taxonomy of terms and agreed upon definitions

- Based on knowledge representation languages
  - RDF (Resource Description Framework)
  - RDFS (RDF Schema)
  - OWL (Web Ontology Language)
What Makes a Technology Semantic?

Makes the Web understandable to computer systems

Has the ability to:
  - Represent knowledge
    - More than just data element definitions
    - Expresses data relationships and process
  - Reason over knowledge to create new knowledge
  - Make connections between data that are non-explicit
  - Deploy a knowledge model for run time consideration
  - Support disparate, distributed resources
    - Ask questions across repositories for integrated results
Semantic Web Languages

- **RDF - Resource Description Framework**
  - General purpose language for representing information on the Web
  - Labels the links on the Web
  - Express relationships between elements
    - Example: Jeep “isTypeof” Ford truck

- **RDFS - RDF Schema**
  - RDF vocabulary description language
  - Describe properties and classes of RDF resources
  - Once relationships are established, reasoning can be performed - via RDFS semantics
Semantic Web Languages

- OWL - Web Ontology Language
  - Greater machine interoperability than RDF/S
    - More expressive power
      - All vs. some
      - Optional vs. required
      - 1-1 vs. 1-many vs. many-1
    - Formal semantics
  - Express meaning of concepts, instances and relationships
Semantic Web Technology Stack

- Self-desc. doc.
- Data
- Rules
- Logic
- Ontology vocabulary
- RDF + rdfschema
- XML + NS + xsmlschema
- Unicode
- URI

Semantic Web LayerCake (Berners-Lee, 99; Swartz-Hendler, 2001)
Where Are We Now?

- RDF, RDFS and OWL are ready for prime time
  - Designs are stable, implementations maturing
- Major research investment translating into application development and commercial spinoffs
  - Oracle to support RDF in database 10.2, OWL in 11.0
  - Adobe embeds RDF in all content
  - IBM SNObase ontology management system
  - HP extending Jena to OWL
  - Cisco, Nokia announcements/use in '05
  - Several new starts in SW space (Cerebra, Siderean, SandPiper)
Adding Business Value

Deploying the Semantic Web
At NASA and the Jet Propulsion Lab

Cassini Mission:
Designing and building spacecraft to go to Saturn
A Fragmented Information Space

- Every project does it differently
  - Unpredictable, inconsistent processes
- No agreed upon best practices for information management
- Document groupings that are separated
  - Search is most often frustrating and unsuccessful
  - Time wasted; decision-making hampered
  - Design and engineering rationales frequently lost for mission teams of the future
JPL Projects Today

Parts Catalogues

Electronic Libraries

E-Mail Archives

Engineering Repositories

Problem Reporting System

Financial Data

Where did I store it?
How do I find it?
Distributed Project Teaming

European Space Agency

NGST

Lockheed

Kennedy

JPL

Goddard

Ball Aerospace

Does Your Information Space Look Like This?
Information Building Blocks

An integrated information architecture made up of several 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, Partner Taxonomies
- Business Rules for data reconciliation
  - You say “tomato”……

- Use new technologies developed for the Semantic Web to enable enhanced capability
  - At this point, mainly RDF
Added Value From the Semantic Web

- Can see all content at once through one interface
  - No need for multiple searches at each repository
  - Not dependent of key word search or file handle

- Can associate engineering documents together that are relevant *no matter where they reside*
  - Design, specification, engineering change requests, waivers, closures, review packages, risk management items, action item notifications and closures
  - View the life cycle of an engineering product from one screen
Unified Search for JPL Projects

User with a query

Data Reconciliation
- Metadata Business Rules
- Schema Translation Models
- Semantic Term Mappings

Flight Project Metadata Catalogue

Mapping to JPL Engineering Taxonomy

JPL Directory
Docushare
PDMS
PFRS
External Partners
Achieving the Vision

Leverage what projects produce in the normal course of their business

- Document trees, matrices
- Document standards, Flight Project Practices processes

- *There are many un-mined sources for semantic processing*
- *What schema already exist in your organization?*
  - Video logs, edit lists, shoot lists, archive inventories, library catalogues, product databases
Hypothetical Advertising Example

User: I am looking for the latest advertising campaign of a particular product so that I can re-use the elements. But I have no idea who worked on it last, what the current version is or where it might be stored.
Integrated Query Results

Marketing Team of Susan X, John W, and Bob Y worked on it

“Oh, and here’s their contact info if I need to ask them questions”

The latest jingle is in the Music Library at this location

Product still shots are in the still store database in Photo Archives

Memo clearing the campaign for use is in the Marketing Library

Footage of the product is in the Video Vault and here is a note
about which footage is to be used by all marketing teams
PhotoStuff - Overview

- Existing annotation toolkits only markup flat text on Web pages
- What about other multimedia on the Web?
- Need a toolkit that allows users to markup multimedia
  - Images, video, audio
Why Semantic Technologies?

- Formal representation language
  - Standards compliant
- Link digital content to existing knowledge
  - Advanced
    - Searching
    - Browsing
- Machine processable content
  - Sharable
Big Picture

Images, photos, graphs, etc. → Ontologies (RDFS/OWL) → RDF/XML (Published image metadata)

http/SSL
Overview

- Load digital media
  - Images and/or videos
  - Web or locally
- Load Semantic Web data
  - Ontologies
  - Instances data/KBs
Overview

- **Annotation**
  - Associate regions with concepts from ontologies
    - Use existing instances
    - Create new instances
- **Publish Annotations**
  - Save the annotations to disk
  - Publish to Web portal
Resources

- W3C’s site:
  - http://www.w3.org/2001/sw/
  - Integrating Applications on the Semantic Web
    - http://www.w3.org/2002/07/swint

- PhotoStuff Homepage:
  - http://www.mindswap.org/2003/PhotoStuff/
Thanks for your Time!

Jayne Dutra
Jayne.E.Dutra@jpl.nasa.gov

Christian Halaschek - Wiener
halasche@cs.umd.edu