

Introduction to Semantic Web

Thomas Baker, Fraunhofer-Gesellschaft

CASHMERE-int Workshop

Göttingen, Germany

2005-02-28



Fraunhofer Gesellschaft

This Talk

- § The Semantic Web Vision

- § Languages of the Semantic Web

 - § Dublin Core to RDF and Ontologies

- § Semantic Web Applications

- § Sources for this talk:

 - § <http://www.w3.org/2004/Talks/1013-semweb-em/talk>

 - § <http://www.w3.org/2004/Talks/0120-semweb-umich/>

The Semantic Web Vision



Fraunhofer Gesellschaft

World Wide Web Consortium

- § Directed by Tim Berners-Lee, “Inventor” of Web
- § Mission: “Lead the Web to its full potential”
- § Hosts: MIT, ERCIM, Keio University
- § Develops Web Standards
 - § HTML, CSS, XML, Security
 - § Web Accessibility Initiative (WAI)
 - § Web Services (SOAP, WSDL, etc.)
 - § Semantic Web languages (RDF, OWL...)

World Wide Web Consortium

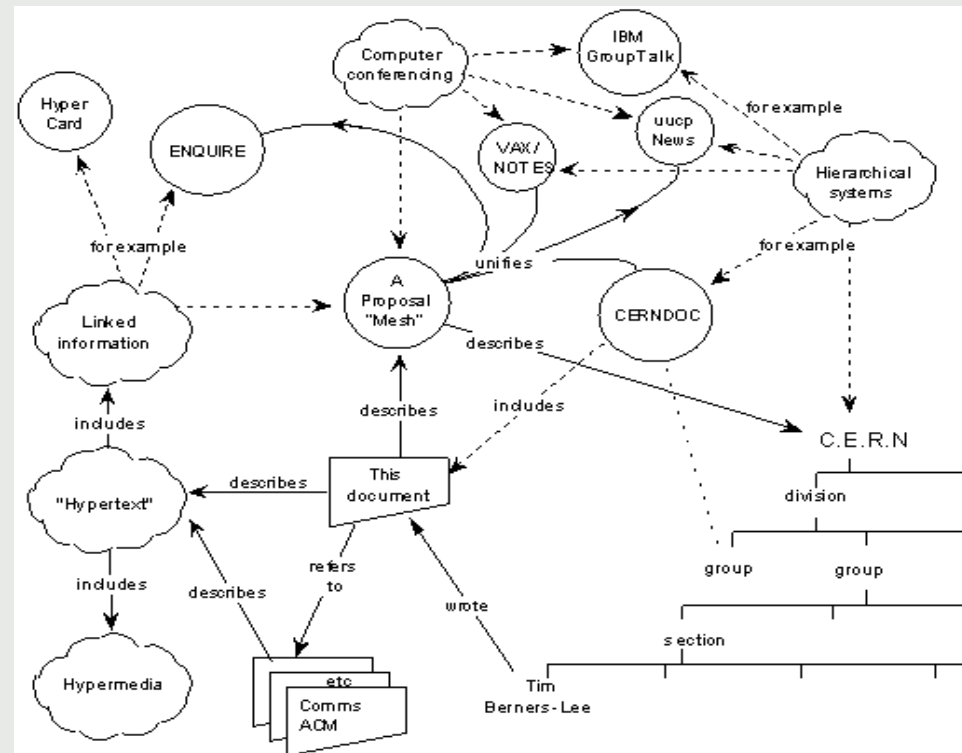
§ Method

- § Technical specifications developed with Working Groups and extensive public review

- § Advanced Development to chart long term architectural directions

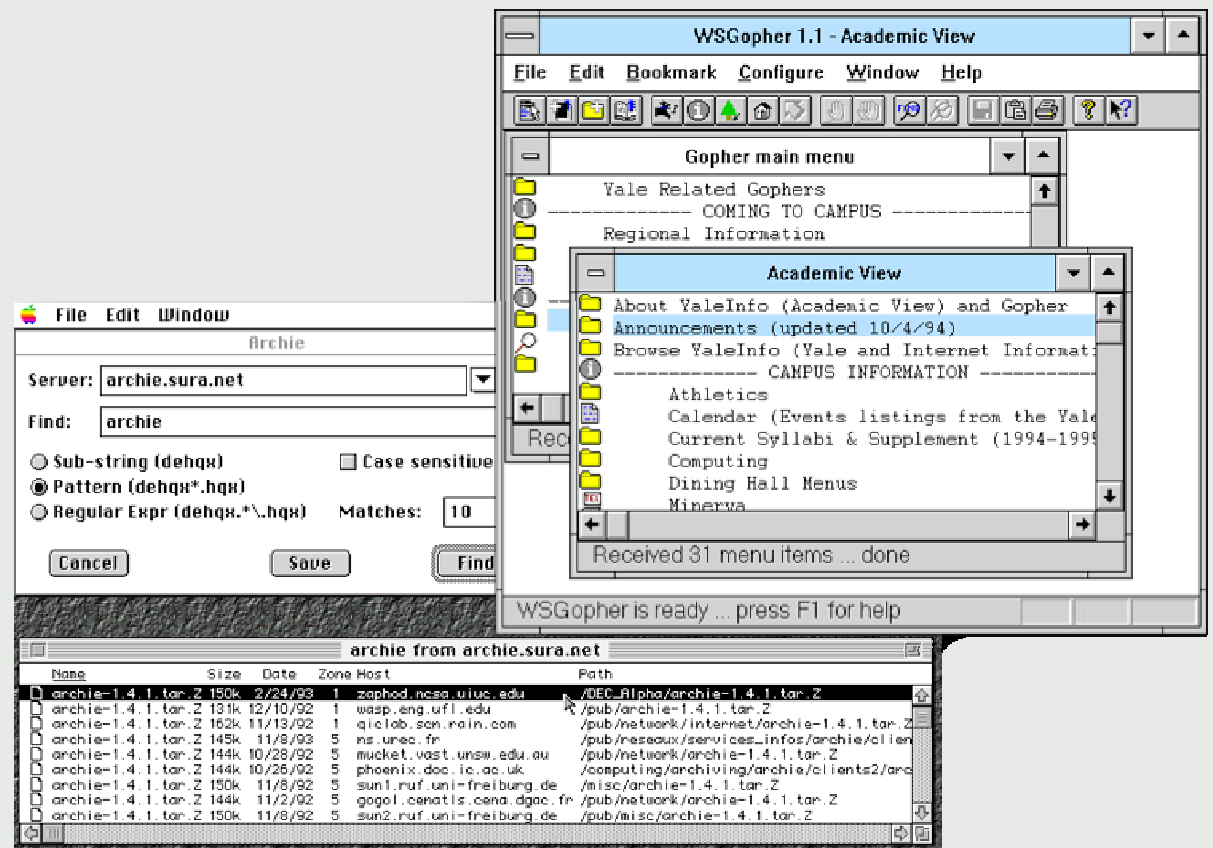
- § Building infrastructure to address technical and social needs of the Web

Semantic Web Vision



Web of Data – the file level

- § Circa 1993
- § FTP, Gopher, Archie:
sharing
resources on
the Internet
- § Stopped at
file level



Web of Data – the text level

§ Circa 1994

§ HTML and URLs

§ Below file level

§ But stopped at text level

The screenshot shows the W3C (World Wide Web Consortium) website. The header includes the W3C logo and the tagline "Leading the Web to Its Full Potential...". Below the header are navigation links: Activities, Technical Reports, Site Index, New Visitors, About W3C, Join W3C, and Contact W3C. A paragraph describes the W3C's mission: "The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential. W3C is a forum for information, commerce, communication, and collective understanding. On this page, you'll find W3C news, links to W3C technologies and ways to get involved. New visitors can find help in [Finding Your Way at W3C](#). We encourage you to read the [Prospectus](#) and learn more about W3C."

The main content area is divided into three columns. The left column, titled "W3C A to Z", lists various W3C activities and documents, including Accessibility, Amaya, Annotate, Binary XML, CC/PP, CSS, CSS Validator, Device Independence, DOM, HTML, HTML Tidy, HTML Validator, HTTP, InXML, Internationalization, Jigsaw, Libwww, MathML, Multimodal Interaction, OWL, Patent Policy, PICS, PNG, Privacy and P3P, Quality Assurance (QA), RDF, Semantic Web, SMIL, SOAP, MLP, Style, SVG, and TAG. The middle column, titled "News", features a section for the "W3C Co-Sponsors 26th Internationalization & Unicode Conference" held from 2004-09-02 to 2004-09-10 in San Jose, CA. The right column, titled "Search", includes a Google search bar and links to "Search W3C" and "Search W3C Mailing Lists".

Below the main content area, there is a section titled "Semantic Web" with a sub-header "Technology and Society domain". It describes the Semantic Web as a common framework for sharing and reusing data across applications, enterprises, and community boundaries. It mentions that the Semantic Web is based on the Resource Description Framework (RDF) and integrates various applications using XML for syntax and URIs for naming. A quote from Tim Berners-Lee is included: "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." - Tim Berners-Lee, James Hendler, Ora Lassus, *The Semantic Web*, Scientific American, May 2001.

Below the Semantic Web section, there is a "News and Events" section with several bullet points: "RDF Data Access Use Cases and Requirements Updated 2004-08-04", "Representing Specified Values in OWL 2004-08-03", "Call for Participation: Public Workshop on Semantic Web for Life Sciences 2004-07-28", and "RDF Data Access Working Group Meets to Select Initial Design 2004-07-26".

A blue arrow points from the text "But stopped at text level" to the "Web of Data" section, which is highlighted in the screenshot. Below the arrow, the text "" is visible.

Semantic Web of Data

- § And now
- § XML, RDF, OWL, URI
- § Below file level
- § Below text level
- § At data level

The image shows a screenshot of the W3C Semantic Web Activity page. The page header includes the W3C logo and the text "Technology and Society domain Semantic Web Activity". The main heading is "Semantic Web". Below it, there is a paragraph describing the Semantic Web as a common framework for sharing and reusing data across application, enterprise, and community boundaries. It mentions that it is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming.

Below the paragraph, there is a quote from Tim Berners-Lee, James Hendler, Ora Lassila, and the Semantic Web Scientific American, May 2001: "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

On this page, there are links for: [Activity Statement](#), [Specifications](#), [Publications](#), [Presentations](#), and [Groups](#).

Nearby, there are links for: [Advanced Development](#), [SWAD Europe](#), [Similar](#), [Semantic Web Coordination](#), [RDF](#), [RDF Core](#), [Web Ontology](#), [Best Practices and Deployment](#), [Interest Group](#), and [Developer Tools](#).

Below the links, there is a section titled "News and Events". It contains two bullet points:

- ▶ [RDF Data Access Use Cases and Requirements Updated](#) 2004-08-04, The RDF Data Access Working Group has released an updated Working Draft of [RDF Data Access Use Cases and Requirements](#). The draft suggests how an [RDF](#) query language and data access can be used in the construction of novel, useful Semantic Web applications in areas like Web publishing, information management, transportation and tourism.
- ▶ [Representing Specified Values in OWL](#) 2004-08-03, The Semantic Web Best Practices and Deployment (SWBPD) Working Group has released the First Public Working Draft of [values in OWL: "value partitions" and "value sets"](#). Comments are welcome for representing modified values and collections of values in the [OWL Web Ontology Language](#).

Below the news section, there is a section titled "People". It contains a link for [Public Workshop on Semantic Web for Life Sciences](#) 2004-07-28, Positioning the Semantic Web for Life Sciences to be held in [Paris](#). The workshop will discuss how Semantic Web technologies such as [RDF](#), [OWL](#), [XQuery](#) integrate with [XML](#) and [XSLT](#) to help to manage modern life sciences research, enable information management of therapies.

Below the people section, there is a section titled "News". It contains a link for [RDF Data Access Use Cases and Requirements](#) in initial design at [W3C](#).

Overlaid on the screenshot are three 3D grid-like structures representing organizational, people, and news levels. The "Organizational" structure is green and has a white circle in the top-right cell. The "People" structure is blue and has a white circle in the top-left cell. The "News" structure is pink and has a white circle in the top-right cell. Arrows point from the text "Organizational", "People", and "News" to their respective structures. Additionally, arrows point from the white circles in the "Organizational" and "People" structures to the "News" structure, indicating a flow or relationship between the levels.

Semantic Web Vision

- § Scenario: Someone needs to see a doctor.
 - § Tim Berners-Lee, Scientific American 2001
 - § Handheld devices with “agents” negotiate:
 - § Doctor’s agent knows prescribed treatment
 - § List of doctors, rated at least “very good”, within 20 miles, accepted by insurance plan.
 - § Devices match calendars to fix appointment.
- § Semantic Web challenge: provide language that expresses both data and rules for reasoning about that data

Semantic Web Vision

§ Web of data

- § Integrating multiple sources to draw new conclusions
- § Architecture for describing all kinds of things (items, collections, services, processes, etc.)
- § Effective management and reuse of data at various scales (personal, group, enterprise, community, web)

Semantic Web Vision

§ URI – Foundation of Semantic Web

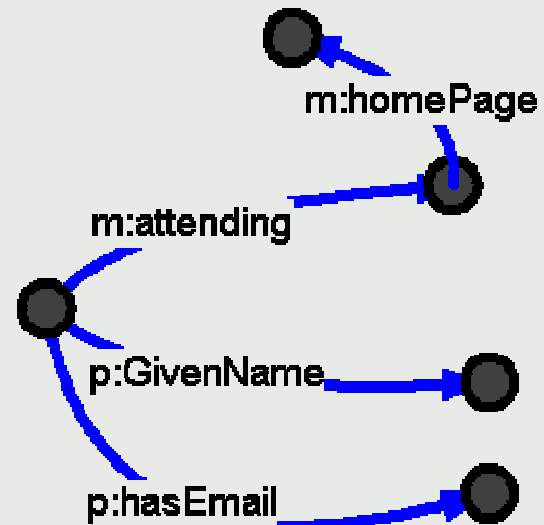
§ Identify **anything** distinctly and uniquely by an opaque string of characters in global URI space

§ physical things – books, cars, people...

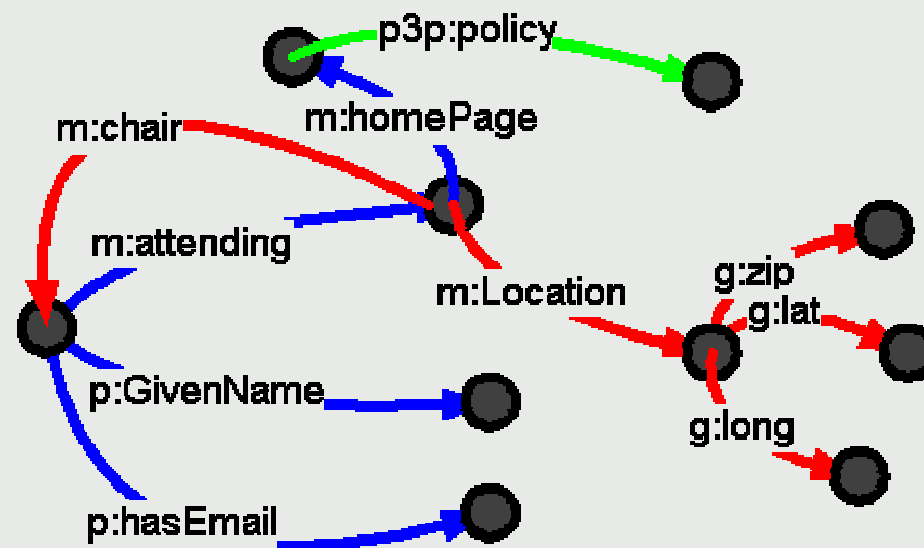
§ digital objects – Web pages, digital images...

§ conceptual things – colors, subjects, metadata terms!

URIs as anchors for merging data



URIs as anchors for merging data



Languages of the Semantic Web



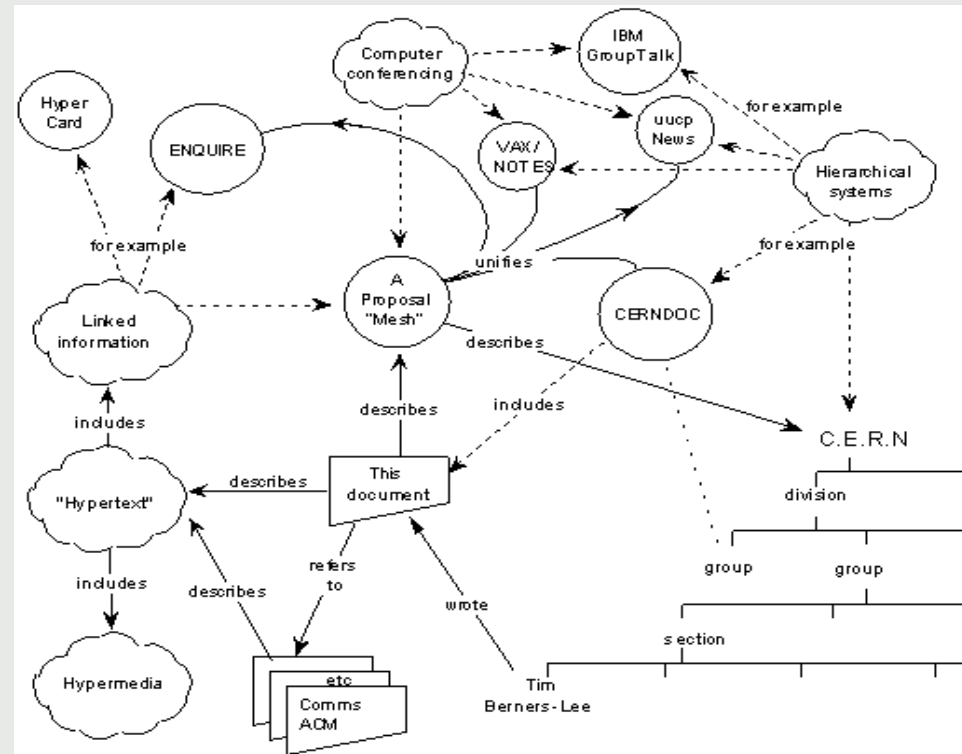
Fraunhofer Gesellschaft

Simple Resource Description

- § Mid-1990s: Web takes off – flood of information
 - § “How can we describe this simply?”
- § “Dublin Core”: Title, Subject, Date...
 - § Pidgin-like “metadata” for describing Web pages
- § Library Science meets Computer Science
 - § Not enough to agree on words – also need grammar!
 - § Machine processing requires agreement on a simple data model

Simple Linked Data Model

Machine-processable statements about **things** (Web pages, organisations, people, concepts, products, etc) and the **links** between them



“Resource Description Framework”

§ 1997: Define metadata vocabularies (like Dublin Core) and use them to make statements

§ **Resource**: anything identifiable with a URI

§ **Description**: statements about properties of resources

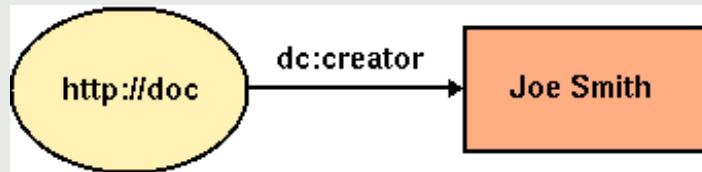
§ **Framework**: a common model for statements using diverse vocabularies

RDF “triples”

- § A simple model for “statements”
 - § Subject: what the statement is about
 - § Predicate: a property of the subject
 - § Object: the value of the property
- § “A natural way to describe the vast majority of the data processed by machines”



Statement about a document using Dublin Core in RDF/XML



```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/" >
  <rdf:Description about="http://docs" >
    <dc:creator> Joe Smith </dc:creator>
  </rdf:Description>
</rdf:RDF>
```

URIs to identify properties

§ Connecting information to its definitions and its context

§ Dublin Core element Creator is denoted by a URI

§ <http://purl.org/dc/elements/1.1/creator>

§ Not just any concept of “color”, but one defined in an identifiable context:

§ <http://pantone.example.com/2004/std#color>

Statements defining the Dublin Core property itself on the Web

```
<rdf:Property rdf:about="http://purl.org/dc/elements/1.1/creator">
<rdfs:label xml:lang="en-US">Creator</rdfs:label>
<rdfs:comment xml:lang="en-US">An entity primarily responsible for
    making the content of the resource.</rdfs:comment>
<dc:description xml:lang="en-US">Examples of a Creator include a
    person, an organisation, or a service. Typically, the name of
    a Creator should be used to indicate the entity.</dc:description>
<rdfs:isDefinedBy rdf:resource="http://purl.org/dc/elements/1.1/"/>
<dcterms:issued>1999-07-02</dcterms:issued>
<dcterms:modified>2002-10-04</dcterms:modified>
</rdf:Property>
```

Semantic Web languages today

§ RDF – Resource Description Framework

§ See <http://www.w3.org/TR/rdf-primer/>

§ OWL Web Ontology Language

§ See <http://www.w3.org/TR/owl-features/>

§ Core specifications are W3C

Recommendations as of February 2004:

§ See <http://www.w3.org/2001/sw/>

Challenge: Independent vocabularies working together

- § Need to formally declare how vocabularies relate to each other
 - § “MARC:illustrator refines DC:contributor”
 - § “RRS:title refines DC:title”
 - § Relationships can be declared in RDF on Web
- § Will require well-understood conventions for declaring and managing vocabularies

Defining Best Practice for Vocabulary Management

- § Identify terms with URI references
- § Articulate maintenance policies for your vocabulary (semantic stability, persistence)
- § Identify historical versions
- § Declare your terms using a formal, machine-processable schema language

Semantic Web Applications



Fraunhofer Gesellschaft

Evolution towards a Semantic Web

- § Not revolution
- § Emphasis on methods for extracting existing data from documents, servers and databases
- § Translate into common form using Semantic Web languages



D2RQ
DBVIEW



XSLT
GRDDL
Schema Annotation



XSLT
GRDDL

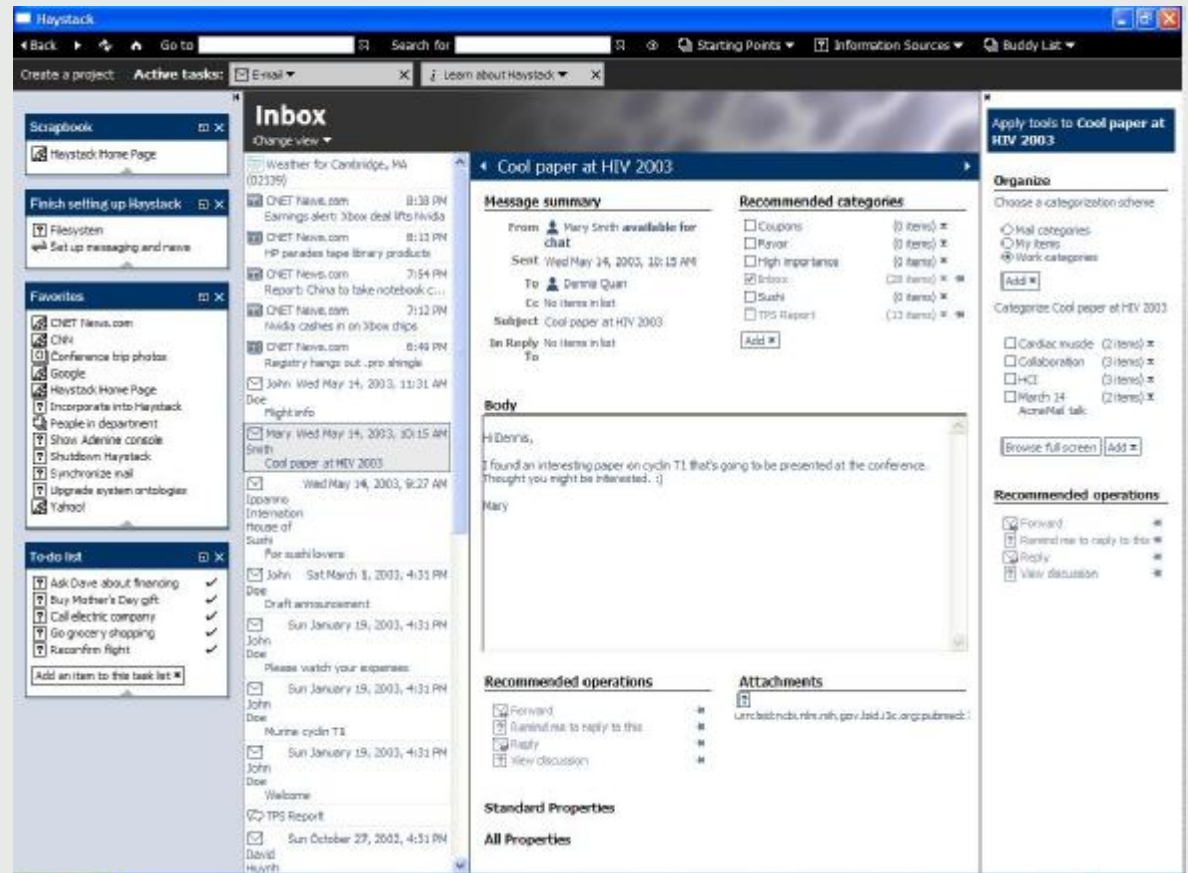


Joseki
DAWG



Personal Information Environments

- § Haystack - User configurable universal information client
 - § benefits from RDF's universal information model
- § Uses RDF for personalization, data, layout, preferences, etc.



<http://haystack.lcs.mit.edu/>

Integrating Enterprise Data

- § Tucana - Enterprise Information Integration
- § Expose diverse data sources as RDF
- § Scalable back-end storage

The screenshot shows the Tucana Technologies Inc. website. The header includes the company logo and navigation links: About Us, Solutions, Customer Relations, Resources, and Partners. The main content area features a large circular image of three people, with the text "Information Fusion... Tucana" next to it. Below this, there are sections for "EII Benefits" (Directed Graph Reasoning, Knowledge Discovery), "Read our FAQ's", and "W3C" and "RDF" logos. The footer contains copyright information and a note about the site design by Style Chocolat.

Tucana TECHNOLOGIES INC. Enterprise Information Integration

About Us Solutions Customer Relations Resources Partners

Tucana Events

Web Seminar
EII - Pull It All Together
Wednesday September 22, 2004

Order Tucana's presentation at the May 2004 World Wide Web Consortium by sending an email to sales@tucanatech.com.

Visit the Resource Center
Get our White Paper "A New Type of Data Management"

EII Benefits

- Directed Graph Reasoning
- Knowledge Discovery

Read our FAQ's

W3C® RDF

With Tucana Information Management Suite at the core of your Enterprise Information Integration (EII) strategy you bring all the power of enterprise knowledge together and put it in the hands of your engineers, scientists, bankers, salespeople or managers.

You **discover the knowledge** you need to

- Outsmart your competitors,
- Increase productivity and creativity and
- Tear down the barriers that stifle innovation.

And best of all you dramatically **reduce the cost and complexity of data management** and **reduce the cost and complexity of rewriting applications** to accommodate business changes.

Tucana's solution gathers stores and analyzes data from relational databases, portals, emails, documents - any or all data sources. Tucana Information Management solution is fast, scalable, agile and standards based - utilizing the technologies of the semantic web.

© Tucana Technologies, Inc. 2004. All rights reserved. Site Design by Style Chocolat

<http://www.tucanatech.com/>

Aggregating Knowledge Bases

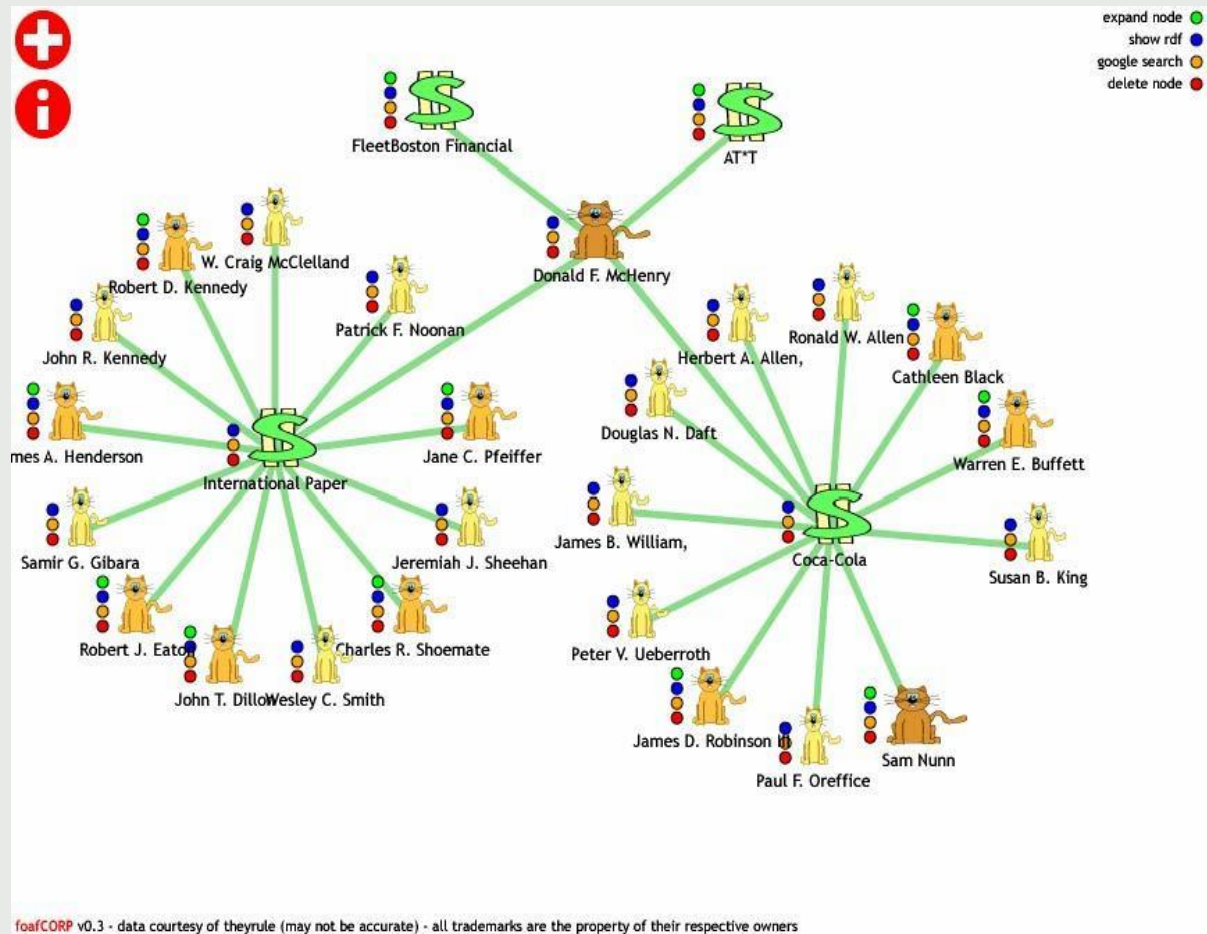
- § TAP
- § Simple tools that treats the Web a giant distributed database.
- § Local, independently managed knowledge bases can be aggregated

The screenshot displays a web browser window with search results for the term 'Sting'. The main content area lists several search results, including the official Sting website, a domain holding page, and a fan club. To the right of the main results, there are several smaller, specialized links or widgets, such as a 'People Magazine Profile' featuring a photo of Sting, 'Information from AllMusic', 'Playing at Tower Records' with a list of albums and prices, 'Buy from CDNow', and a link to 'sting.sdsu.edu'. The browser's address bar at the bottom shows the URL 'http://tap.stanford.edu/'.

<http://tap.stanford.edu/>

Visualizing Social Networks

§ FoafCorp:
visualizing
corporate
boards of
directors



<http://rdfweb.org/foafcorp/intro.html>

Web'ifying Thesauri

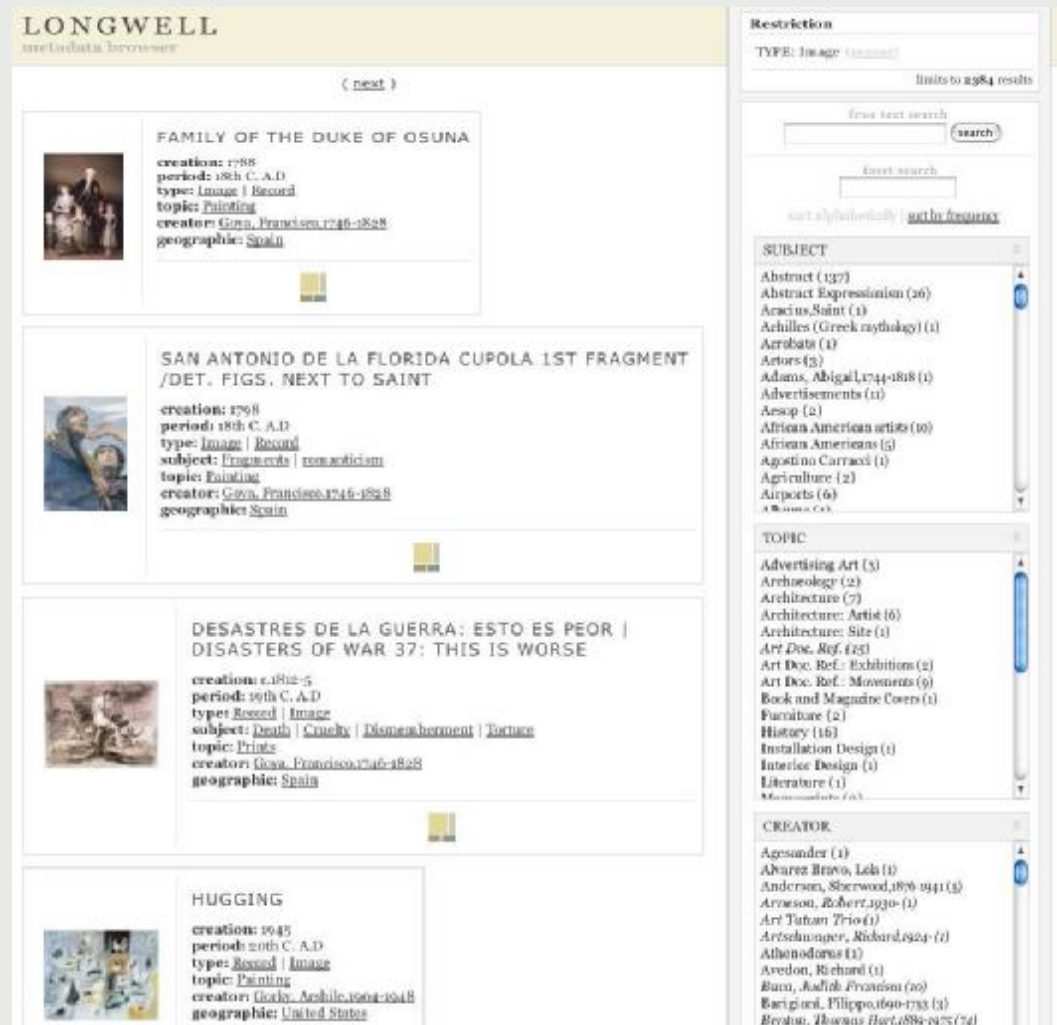
- § SKOS – an RDF vocabulary for thesauri
 - § Encoding and mapping of thesauri, controlled vocabularies
 - § Bridges library classification and Web technologies
 - § <http://www.w3.org/2004/02/skos/core/guide/>
- § A product of SWAD-Europe - Supporting Semantic Web standards in Europe
 - § Targeted research, software applications, outreach
 - § Topics: calendaring, social networks, images, geospatial, internationalization, RDF storage, etc.

Bridging Institutional Repositories

- § Project Simile - Semantic Web meets Digital libraries and personal information management
 - § Partners: W3C, HP, MIT Libraries, MIT CSAIL
- § Implement a digital asset management architecture based on Web standards
 - § Add useful “views” to a digital artifact and bind those views to consuming services
- § Leverage and extend DSPACE, enhancing its support for managing heterogeneous data

“RDF Browsers”

- § Common interface, framework for navigation
- § Architecture supports integration of heterogeneous data sources
- § Tools for exposing content collections in RDF
- § <http://simile.mit.edu/>



Integrating Life Science Data

- § Connecting information: gene, diseases, cures
- § Scientists in different locations, working on different problems, integrating results into coherent whole
- § Recognized need for effective data integration from heterogeneous collections
- § Increasingly available datasets in RDF
- § Increasing scientific / vendor interest
 - § Semantic Web and Life Sciences Workshop, Oct 27-28, Cambridge MA, <http://www.w3.org/2004/07/swls-ws.html>

W3C Semantic Web Activity

Phase 2

- § Feb 2004: Core specs are W3C Recommendations
 - § Open Standards and Open Source tools, technologies for modeling real world resources; sharing these models across the Web.
- § Mar 2004: Phase 2 launched
 - § RDF Data Access - “Joining the Web”
 - § Best Practices and Deployment
 - § Advanced Development
- § Deployment / Facilitating 'Network Effect'

W3C Semantic Web Activity

RDF Data Access (DAWG)

§ To define an HTTP and/or SOAP protocol for selecting instances of RDF

§ ‘Join’ Web data as easily as merging tables in a local relational database.

§ Use Cases

§ Personal Information Management, transportation, tourism, product life-cycle data management, publishing...

§ Outputs

§ Use Cases and Requirements - Aug 2004

§ SPARQL - Query Language Specification - Oct 2004

W3C Semantic Web Activity

Best Practices and Deployment

- § To provide guidance for developers of Semantic Web applications.
- § Best Practice notes for ontology engineering, vocabulary development...
- § Educational material and demo applications.
- § Support transformation of ontologies and thesauri to RDF/OWL
- § <http://www.w3.org/2001/sw/BestPractices/>

W3C Semantic Web Activity

Advanced Development

§ Collaborative development

§ Creation of core components (e.g. libwww) that will form the basis for the Semantic Web.

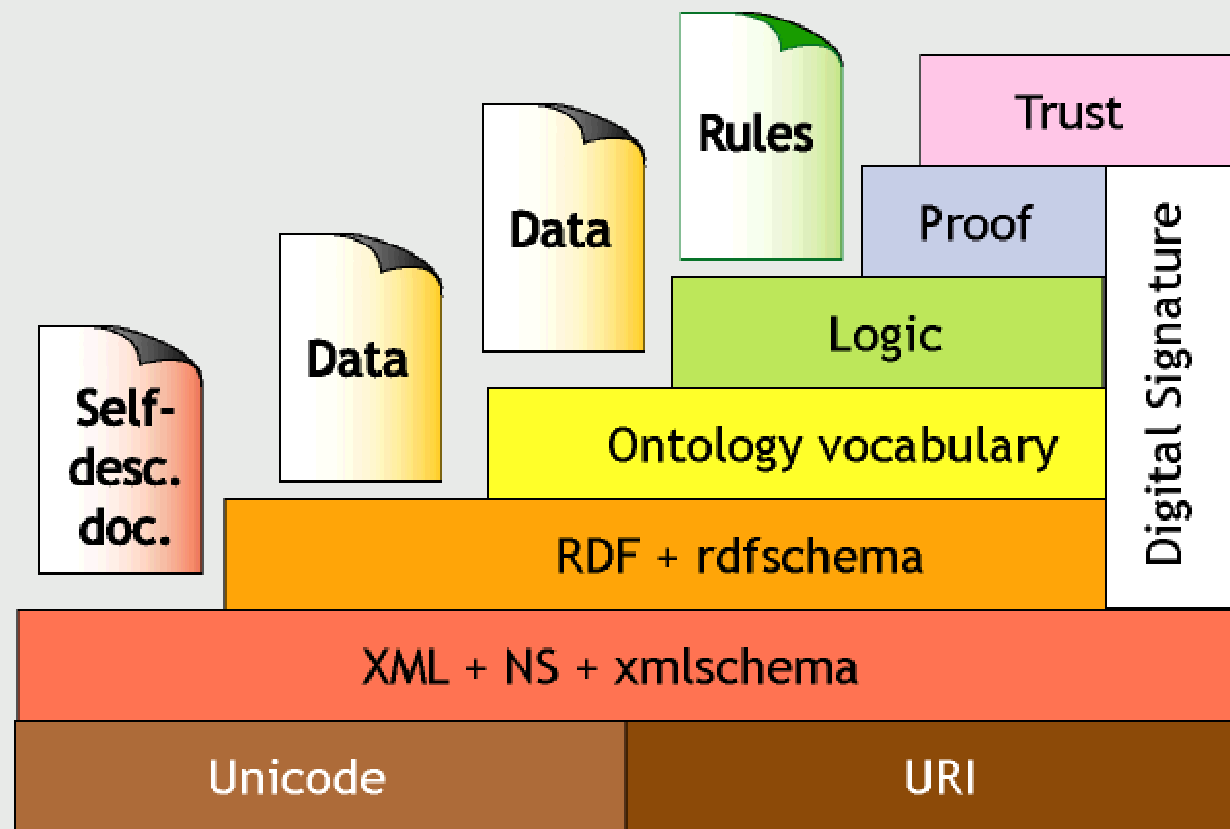
§ Facilitate Semantic Web deployment and identify futures areas of standardization

§ <http://www.w3.org/2000/01/sw/>

A Common Framework

- § Many organizations are starting to realize they need ‘digital libraries’
 - § Even if they don’t call it that
- § Common data representation
 - § Common description framework and architecture reduces (technical / social) costs and is more efficient
 - § Everyone benefits

Towards the technology of trust



Conclusions

- § Core Specifications in place
- § More applications / toolkits / software every day
- § A new class of Semantic Web applications at individual, enterprise, and Web scale
- § Semantic Web based on a new type of human and machine language

Additional Information

- § W3C World Wide Web Consortium
 - § <http://www.w3.org>
- § Semantic Web Initiative Home Page
 - § <http://www.w3.org/2001/sw/>
- § Dublin Core Metadata Initiative
 - § <http://dublincore.org/>
- § Eric Miller, W3C Semantic Web Activity Lead
 - § <http://www.w3.org/People/EM/>
- § W3C Semantic Web Best Practices and Deployment Working Group
 - § <http://www.w3.org/2001/sw/BestPractices/>

thomas.baker@izb.fraunhofer.de