

World Wide Web Consoritum Semantic Web and Opportunities for the Oil and Gas Industry

Steve Bratt (mailto:steve@w3.org)
Chief Executive Officer
World Wide Web Consortium (http://www.w3.org/)
January 2008

This talk:

http://www.w3.org/2008/Talks/0130-bratt-W3C-Energy/W3C-SemWeb-Appsp.pdf

Outline

- World Wide Web Consortium
 - Mission, organization
 - Summary of technologies
- Semantic Web
 - One slide tutorial
 - Market trends
- Semantic Web Applications
 - Rapid rise in work in the field
 - W3C's Health Care / Life Sciences initiative
- Opportunities for the Oil and Gas Industry
 - How to move forward
 - Discussion



World Wide Web Consortium

- Mission, organization
- Summary of technologies

The Leading Web Standards Orgnization

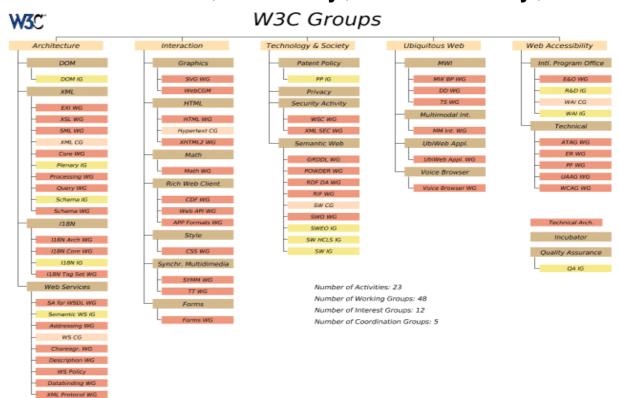
- W3C engineers the Web's foundation
 - 1000+ technologists in 60 groups

 (x)HTML, XML, CSS, VoiceXML, Mobility, Ubiquity, Video, Web Services, Web 2.0, Semantic Web, Security, Accessibility,

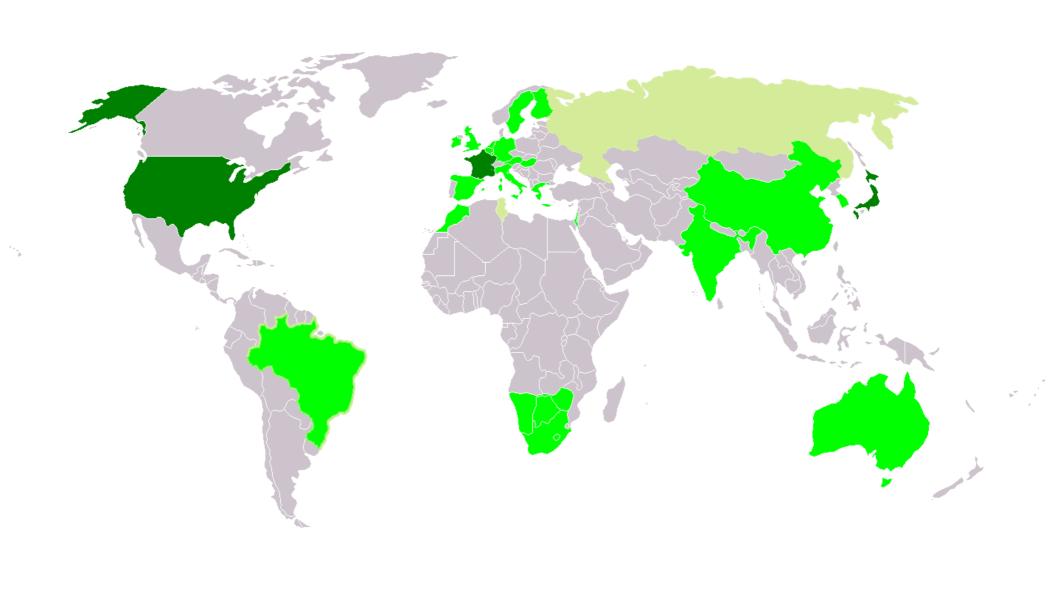
Privacy, I18N,

Architecture

- 430 Members
- 60 Technical Staff
- 40+ Liaisons



Offices in 21 Countries (more under consideration)





Semantic Web

- One-slide tutorial
- Market trends
- Applications

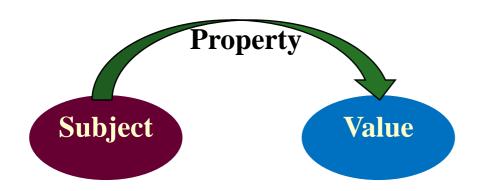
Semantic Web Introduction, Tutorial, Applications

- http://www.w3.org/People/Ivan/CorePresentations/
- ... Compiled by Ivan Herman, W3C Semantic Web Activity Lead



Semantic Web @ W3C

- Web 1.0 = Linked Documents
- Semantic Web = Linked Data (Web 3.0*)



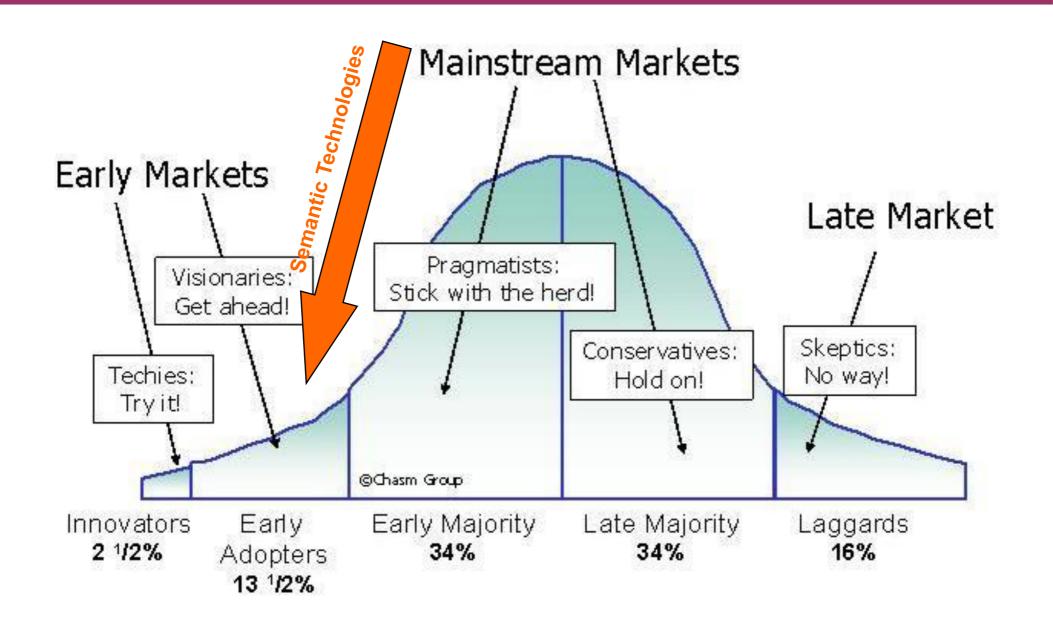
... where Subjects, Properties, Values can each have their own URI, and thus are universally unique and linkable across the Web

- Web becomes a global, relational database
- Semantic Web Activity @ W3C
 - URIs, <u>RDF</u> (above), <u>Web Ontology Language</u>
 - Query, Rules, Content Labeling, Case Studies and Use Cases



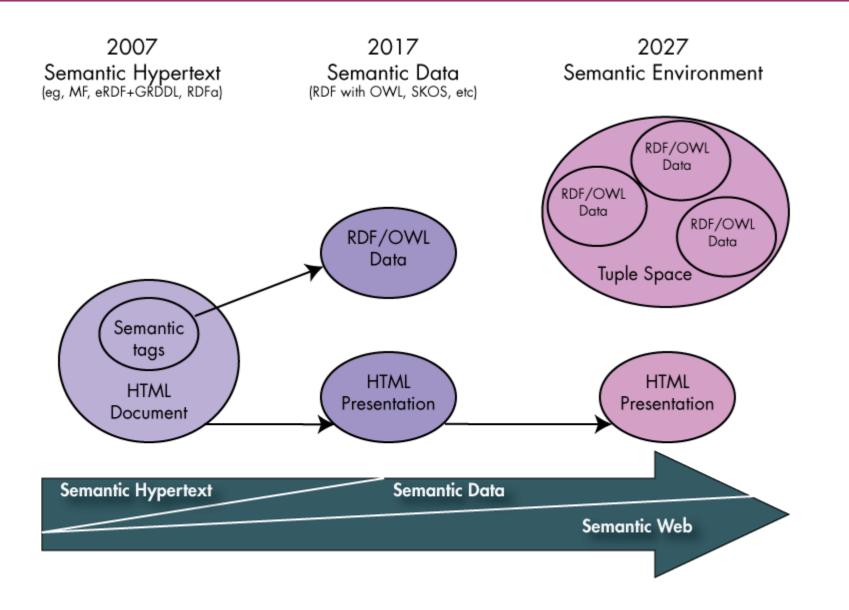
^{*}New York Times, InternetNews

Technology Adoption Life Cycle





The 2007 Gartner Predictions



See extra slides at end regarding how Semantic Web may grow, and a C-level guide to getting started.

Based on: "Finding and Exploiting Value in Semantic Web Technologies on the Web", Gartner Research Report, May 2007



What We Are Seeing

Domains such as

- digital libraries, defense, eGovernment, energy sector, financial services, health care, life sciences, oil and gas industry ...
- ... are showing interest and organizing themselves.
- Organizations with ...
 - complex data management and challenges
 - intellectual interest and resources
- ... are starting to employ Semantic technologies.
- Major companies offer (or will offer) Semantic Web tools, using Semantic technologies and often participating in W3C:
 - Adobe, Oracle, IBM, HP, Software AG, GE, Northrop Gruman, Altova, Dow Jones, Novartis, Boeing, Pfizer, Telefónica, ILOG, HP, Agfa, SRI International, Fair Isaac Corp., Oracle, Boeing, IBM, Chevron, Siemens, Nokia, Merck, Pfizer, AstraZeneca, Sun, Eli Lilly, ...



Semantic Web Education and Outreach Group

At present there are

- 16 case studies and 10 use cases (October 2007)
- from 11 different countries around the globe
- activity areas of submitters include: automotive, broadcasting, financial institution, health care, oil & gas industry, pharmaceutical, public and governmental institutions, publishing, telecommunications, ...
- usage areas include: data integration, portals with improved local search, business organization, B2B integration, ...
- Remember this URI:

http://www.w3.org/2001/sw/sweo/public/UseCases/



Survey of Semantic Technologies Used in the Oil and Gas Industry

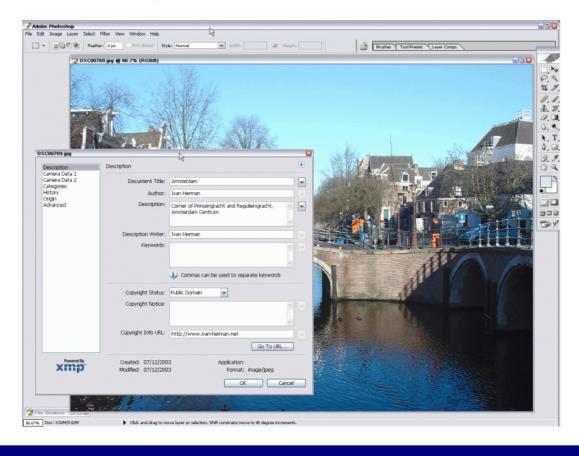
- Fluor's Accelerating Deployment of ISO 15926 project targets integration, exchange, and hand-over of information in process industries during the life cycle of a plant.
- InfoWeb, a plant data specialist in the Netherlands, leverages the Semantic Web for developing the ISO 15926 knowledge base
- The Norwegian Daily Production Report (DPR) project is standardizing production data reporting, making it more accessible for authorities and license partners.
- The <u>Active Knowledge Systems for Integrated Operations</u> project is developing an integrated system in historical and realtime data and knowledge management to support drilling operations.
- The Integrated Information Platform project sponsored by the Norwegian Research Council, aims to create an information platform for the industry by integrating ontologies from several industrial data and technology standards and also by creating new ontologies.

Courtesy of Frank Chum, Chevron (SWEO Use Case)



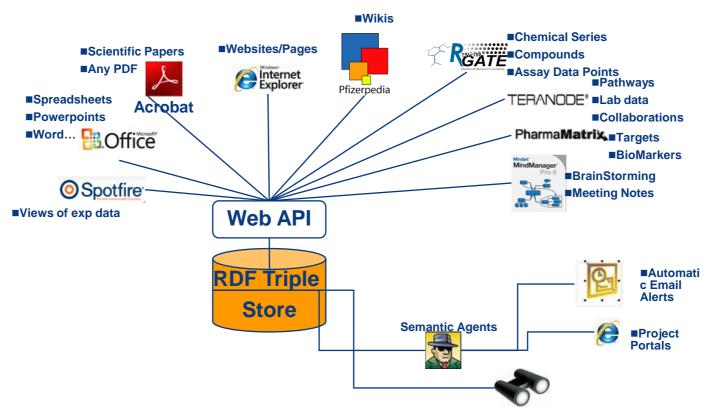
Adobe's XMP

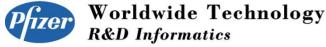
- Metadata is added by, e.g., Photoshop into files in RDF
- XMP is a way of embedding + vocabulary + a set of (public) tools (there are also 3rd party tools to extract the RDF content)
- Used by a number of platform solutions



Ontology Controlled Annotation

 Annotation of different data formats all along the full drug discovery process...

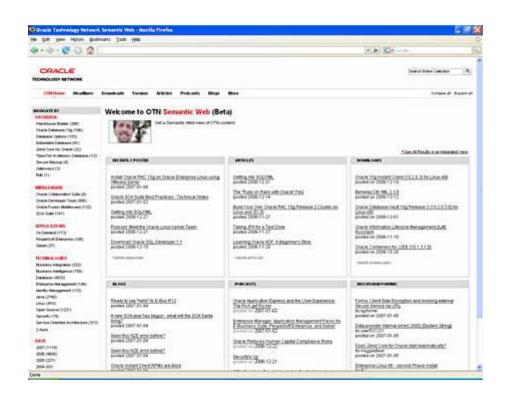




Courtesy of Giles Day, Pfizer

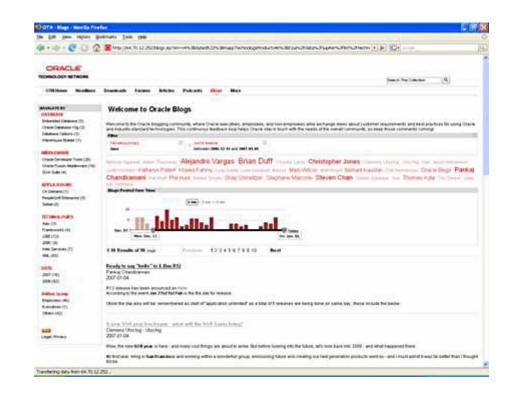


Oracle's Technology Network Portal



Re-group, categorize, etc content (using a taxonomy)

Aggregates many source of content

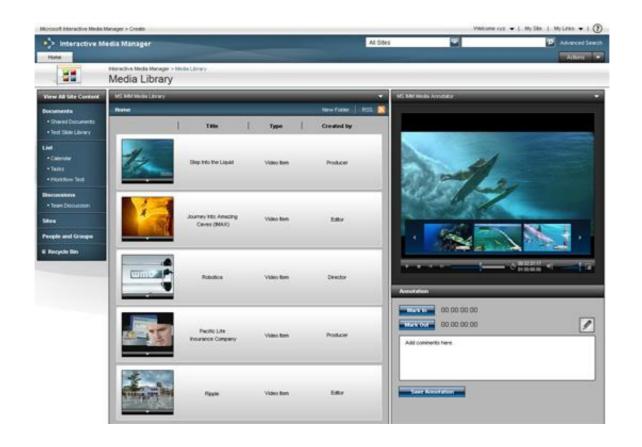


Courtesy of Mike DiLascio, Siderean Software, and Justin Kestelyn, Oracle Corporation (SWEO Case Study)



Microsoft Vista's Interactive Media Manager

- Uses an RDF/SPARQL/OWL based metadata framework
 - eg, for a better control over relationships among media assets and categories
- Custom OWL ontologies can be created and imported



Repair and Diagnostics Documentation

- Store repair and diagnostic operations in a repository with an OWL model
- A "diagnostic engine" generates the manuals on the fly, using RDF for information exchange

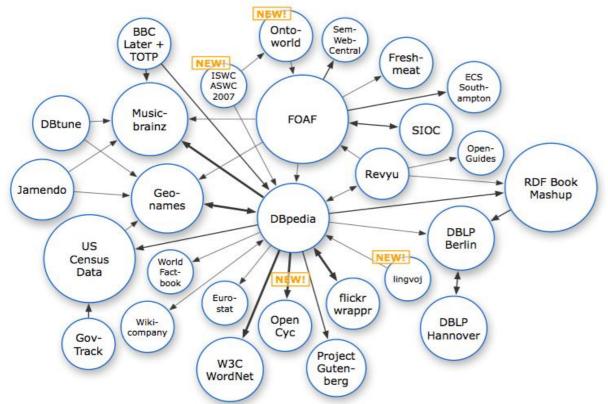


Courtesy of François-Paul Servant, Renault, (SWEO Use Case)



Linking Open Data Project

- Goal: "expose" open datasets in RDF
- Set RDF links among the data items from different datasets
- Set up SPARQL endpoints to query the data, too
- Over 2 billion triples. 3 million "links" (November 2007)



http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData



Extracting Structured Data from Wikipedia

```
http://en.wikipedia.org/wiki/Calgary
```



Courtesy of Chris Bizer, Free University of Berlin



Automatic Links Among Open Datasets

```
<http://dbpedia.org/resource/Calgary>
                                                       DBpedia
 owl.sameAs <http://sws.geonames.org/5913490>;
              <http://sws.geonames.org/5913490>
                owl:sameAs <http://DBpedia.org/resource/Calgary>
                wgs84_pos:lat "51.050112282";
                wgs84 pos:long "-114.085285152";
Geonames
                 sws:population "968460"
```

Processors can switch automatically from one to the other...



Deployment @ W3C

Increasing focus on end users

- Supports application of standards to real, important problem
- Improves standards: Use cases,
 requirements, implementation, testing



- "use of Semantic Web technologies ...
 to improve collaboration, research and
 development, and innovation adoption"
- Agfa, AstraZeneca, Cleveland Clinic, Eli Lilly, HL7, Merck,
 Partners, Pfizer (60+ participants from 40 organizations)



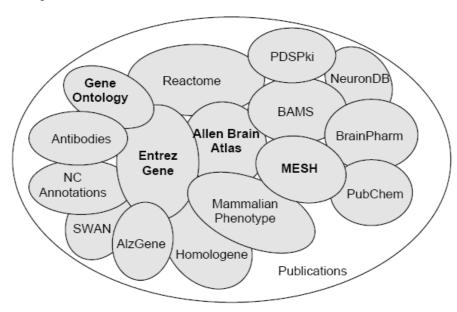
Sample HCLS Task Forces

- BioRDF. Expose biomedical data to RDF/OWL
- BIONT. Support ontology needs for other groups
- Drug Safety and Efficacy. Support monitoring drug safety, pharmacovigilance.
- Adaptable Clinical Protocols and Pathways. Represent and compute applicability of protocols to dynamically changing patient status
- Clinical Observations Interoperability. Establish new collaboration with health care industry players, and work on issues at intersection of electronic medical records and health care org. needs.



Integrating Key Datasets

- W3C HCLS IG has already exposed a number of public datasets and ontologies in RDF and OWL
 - assign URI-s to bio entities
 - data converted or made reachable in RDF
 - use reasoners to infer extra triples to increase expressiveness
 - query the data with SPARQL and visualization tools
 - around 400M triples so far...





Use SPARQL to Query ...

"find me genes involved in signal transduction that are related to pyramidal neurons"

```
prefix ao: <a href="http://purl.org/obo/owl/GO#>">http://purl.org/obo/owl/GO#>">
prefix rdfs: <a href="mailto:rdf">rdf</a>: <a href="mailto:rdf">rdf</a>-schema#>
prefix owl: <a href="mailto:ref">http://www.w3.org/2002/07/owl#></a>
prefix mesh: <a href="http://purl.org/commons/record/mesh/">http://purl.org/commons/record/mesh/></a>
prefix sc: <a href="http://purl.org/science/owl/sciencecommons/">http://purl.org/science/owl/sciencecommons/</a>>
prefix ro: <a href="http://www.obofoundry.org/ro/ro.owl#>">http://www.obofoundry.org/ro/ro.owl#>">http://www.obofoundry.org/ro/ro.owl#>">
                                                                                                          Mesh: Pyramidal Neurons
select ?genename ?processname
where
{ graph <a href="mailto:rg/commons/hcls/pubmesh">graph <a href="mailto:rhttp://purl.org/commons/hcls/pubmesh">http://purl.org/commons/hcls/pubmesh</a>
    ?paper ?p mesh:D017966 .
    ?article sc:identified by pmid?paper_
    ?gene sc:describes_gene_or_gene_product_mentioned_by ?article.
                                                                                                           Pubmed: Journal Articles
 graph <a href="mailto:rg/commons/hcls/goa">http://purl.org/commons/hcls/goa>
    ?protein rdfs:subClassOf ?res.
     ?res owl:onProperty ro:has_function.
    ?res owl:someValuesFrom ?res2.
    ?res2 owl:onProperty ro:realized as.
    ?res2 owl:someValuesFrom ?process.
 graph <a href="mailto:rg/commons/hcls/20070416/classrelations">relations</a>
   {{?process < http://purl.org/obo/owl/obo#part of > qo:GO 0007166}
                                                                                                          Entrez Gene: Genes
    {?process rdfs:subClassOf go:GO 0007166 }}
     ?protein rdfs:subClassOf ?parent.
    ?parent owl:equivalentClass ?res3.
    ?res3 owl:hasValue ?gene.
 graph <a href="mailto:rg/commons/hcls/gene">http://purl.org/commons/hcls/gene</a>
  { ?gene rdfs:label ?genename }
 graph <a href="mailto:rg/commons/hcls/20070416">graph <a href="mailto:rth:rth://purl.org/commons/hcls/20070416">http://purl.org/commons/hcls/20070416</a>
    ?process rdfs:label ?processname}
                                                                                                           GO: Signal Transduction
                                                                  Inference required
```

Courtesy of Susie Stephens, Eli Lilly, Alan Ruttenberg, Science Commons, and the W3C HCLS IG



Results that Make Sense to Experts ...

Many of the genes are indeed related to Alzheimer's Disease through gamma secretase (presenilin) activity

```
adenylate cyclase activation
DRD1, 1812
ADRB2, 154
                         adenylate cyclase activation
ADRB2, 154
                         arrestin mediated desensitization of G-protein coupled receptor protein signaling pathway
DRD11P, 50632
                         dopamine receptor signaling pathway
DRD1, 1812
                         dopamine receptor, adenylate cyclase activating pathway
DRD2, 1813
                         dopamine receptor, adenylate cyclase inhibiting pathway
                         G-protein coupled receptor protein signaling pathway
GRM7, 2917
                         G-protein coupled receptor protein signaling pathway
GNG3, 2785
GNG12, 55970
                         G-protein coupled receptor protein signaling pathway
DRD2, 1813
                         G-protein coupled receptor protein signaling pathway
ADRB2, 154
                         G-protein coupled receptor protein signaling pathway
                         G-protein coupled receptor protein signaling pathway
CALM3, 808
HTR2A, 3356
                         G-protein coupled receptor protein signaling pathway
DRD1, 1812
                         G-protein signaling, coupled to cyclic nucleotide second messenger
SSTR5, 6755G-protein signaling, coupled to cyclic nucleotide second messenger
                         G-protein signaling, coupled to cyclic nucleotide second messenger
MTNR1A, 4543
CNR2. 1269
                         G-protein signaling, coupled to cyclic nucleotide second messenger
HTR6, 3362
                         G-protein signaling, coupled to cyclic nucleotide second messenger
GRIK2, 2898
                         glutamate signaling pathway
GRIN1, 2902
                         glutamate signaling pathway
GRIN2A, 2903
                         ğlutamate siğnaling pathway
GRIN2B, 2904
                         glutamate signaling pathway
                         integrin-mediated signaling pathway
ADAM10, 102
                         negative regulation of adenylate cyclase activity
GRM7, 2917
LRP1, 4035
                         negative regulation of Wnt receptor signaling pathway
ADAM10, 102
                         Notch receptor processing
ASCL1, 429
                         Notch signaling pathway
                         serotoniň receptor signáling pathway
HTR2A, 3356
ADRB2, 154
                         transmembrane receptor protein tyrosine kinase activation (dimerization)
                         transmembrane receptor protein tyrosine kinase signaling pathway
PTPRG, 5793
EPHA4, 2043
                         transmembrane receptor protein tyrosine kinase signaling pathway
NRTN, 4902
                         transmembrane receptor protein tyrosine kinase signaling pathway
CTNND1, 1500
                         Wnt receptor signaling pathway
```

Courtesy of Susie Stephens, Eli Lilly, Alan Ruttenberg, Science Commons, and the W3C HCLS IG



Opportunities for the Oil and Gas Industry

- How to move forward
- Discussion



Potential Value of Working in W3C

Leadership

- Introduce ideas through <u>submissions</u>, <u>workshops</u>, <u>Incubator Groups</u>
- Influence standards => Working Groups, review, implementation

Early insight into market trends

- Access world's top Web technologists from <u>Members</u> and <u>Team</u>
- Plan for emerging technologies & markets through Member access

Productive, Neutral Environment

- People, process, experience, collaboration tools, global scope

Promoting image as innovator

- Participate in int'l media activities, press releases, testimonials
- Display logo on <u>W3C site</u> (300K visits/day) & <u>W3C logo</u> on your site

(Membership / Benefits / How to join W3C / "At a Glance" brochure)



Interest in a Energy Industry Group?

- Chevron has been a W3C Member since 2000. POSC Caesar is joining. Many vendors here.
- Candidate "next steps" for discussion:
 - Engage more domain experts and companies.
 - Develop liaisons with relevant industry organizations
 - Send a "Fellow" to W3C
 - Organize a W3C workshop
 - Create an Interest Group
 - ... to explore, prototype, build community
 - Create an Incubator Group
 - To develop requirements, use cases, draft specifications, ontologies quickly
 - Created a Working Group
 - ... to create Web standards



For more information



http://www.w3.org/

W3C Membership:

http://www.w3.org/Consortium/membership

Thanks to Ivan Herman, W3C, and colleagues in the Semantic Web Education and Outreach group for compiling most of the sample applications presented in these slides.

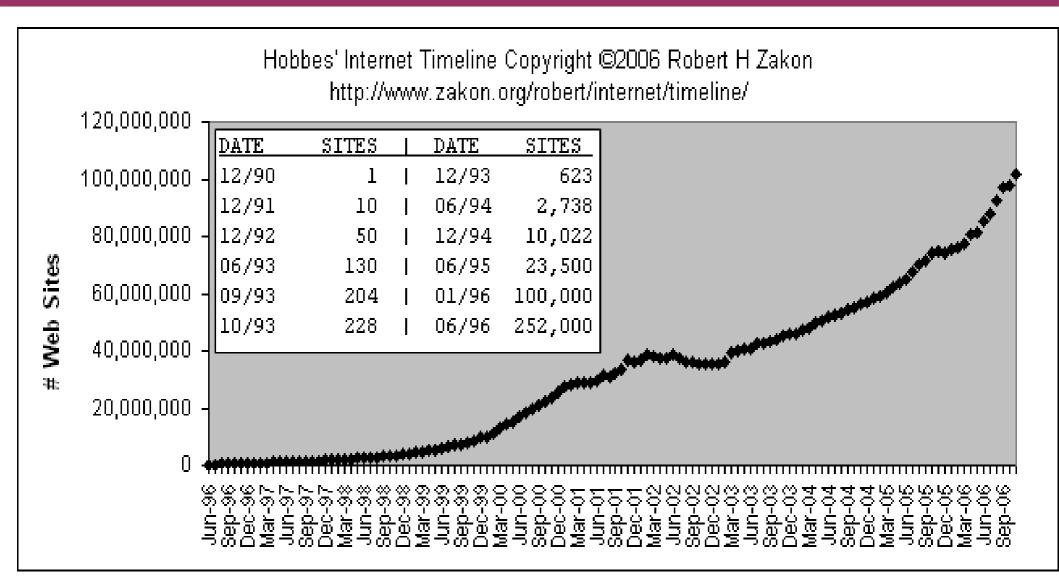


Extra Slides Follow

- Extra slides follow
 - How the Semantic Web is taking off
 - C-level do-s and don't-s for leveraging Semantic technologies
 - Additional examples



Internet Growth Driven by Open Web &Standards



Sources: http://www.zakon.org/robert/internet/timeline/

http://www.internetworldstats.com/stats.htm



May start with small communities

- The needs of a deployment application area:
 - have serious problem or opportunity
 - have the intellectual interest to pick up new things
 - have motivation to fix the problem
 - its data connects to other application areas
 - have an influence as a showcase for others
- The high energy physics community played this role for the Web in the 90's



Some deployment communities

- Some examples: digital libraries, defense, eGovernment, energy sector, financial services, health care, oil and gas industry, life sciences ...
- Health care and life science sector is now very active
 - also at W3C, in the form of an Interest Group



CEO guide for SW: the "DO-s"

- Start small: Test the Semantic Web waters with a pilot project [...] before investing large sums of time and money.
- Check credentials: A lot of systems integrators don't really have the skills to deal with Semantic Web technologies. Get someone who's savvy in semantics.
- Expect training challenges: It often takes people a while to understand the technology. [...]
- **Find an ally**: It can be hard to articulate the potential benefits, so find someone with a problem that can be solved with the Semantic Web and make that person a partner.

Source: <u>BusinessWeek Online</u>, April 2007

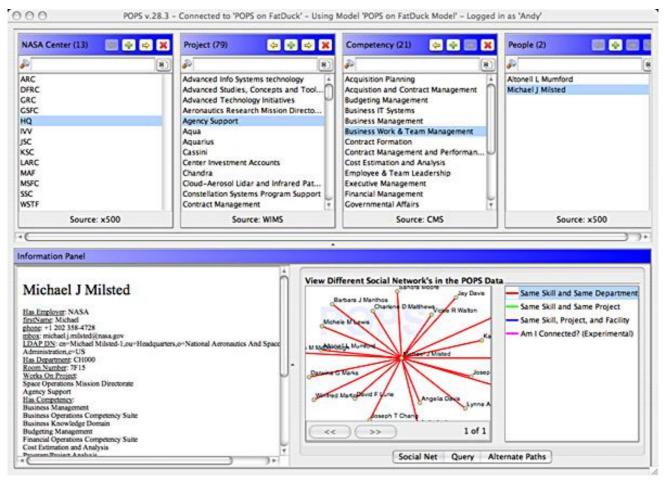
CEO guide for SW: the "DON'T-s"

- Go it alone: The Semantic Web is complex, and it's best to get help.
 [...]
- Forget privacy: Just because you can gather and correlate data about employees doesn't mean you should. Set usage guidelines to safeguard employee privacy.
- Expect perfection: While these technologies will help you find and correlate information more quickly, they're far from perfect. Nothing can help if data are unreliable in the first place.
- Be impatient: One early adopter at NASA says that the potential benefits can justify the investments in time, money, and resources, but there must be a multi-year commitment to have any hope of success

Source: <u>BusinessWeek Online</u>, April 2007

Find the right experts at NASA

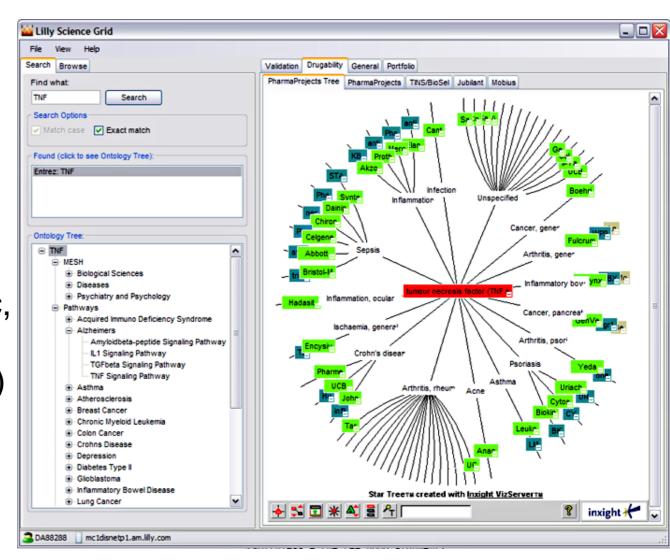
 Expertise locater for nearly 20,000 NASA civil servants using RDF integration techniques over 6 or 7 geographically distributed databases, data sources, and web services...



Courtesy of Kendall Clark, Clark & Parsia, LLC

Eli Lilly's Target Assessment Tool

- Better prioritization of possible drug target, integrating data from different sources and formats
- Integration, search, etc, via ontologies (proprietary and public)

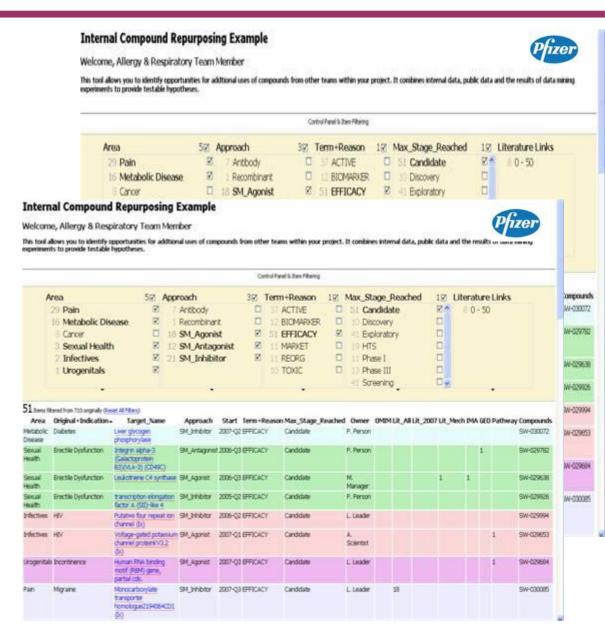


Courtesy of Susie Stephens, Eli Lilly (SWEO Case Study)



Applications may be relatively simple

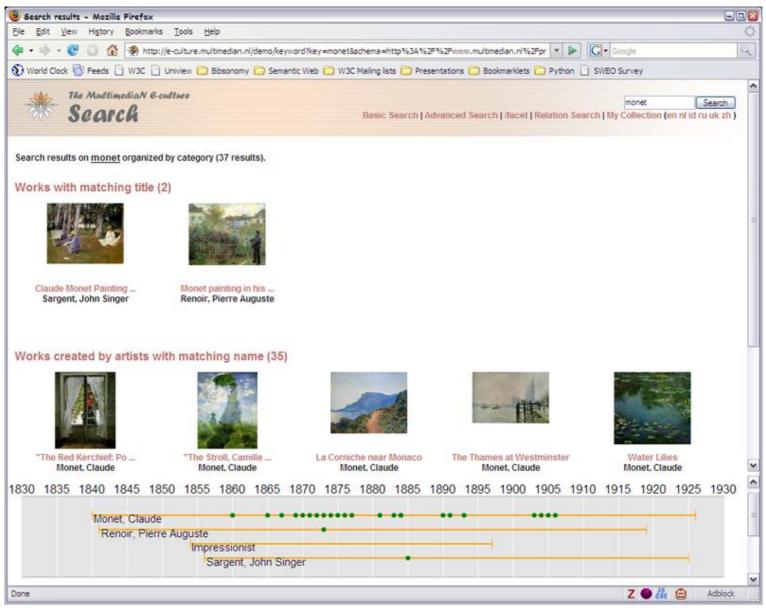
- Goal: reuse of older experimental data
- Keep data in databases or XML, just export key "facts" as RDF
- Use a faceted browser to visualize and interact with the result



Courtesy of Nigel Wilkinson, Lee Harland, Pfizer Ltd, Melliyal Annamalai, Oracle (SWEO Case Study)



Semantic portal for art collections

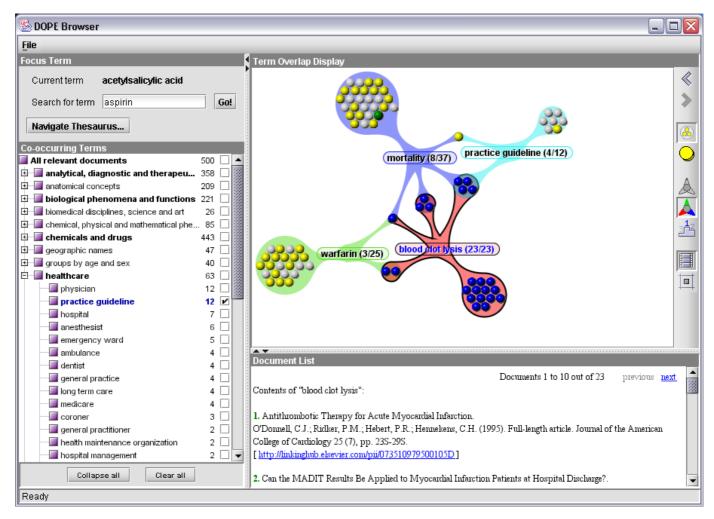


Courtesy of Jacco van Ossenbruggen, CWI, and Guus Schreiber, VU Amsterdam



Elsevier's DOPE browser

- Single interface to multiple data sources (in life sciences)
- Integration, search, etc, via thesauri and metadata in RDF(S)

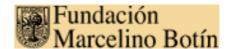


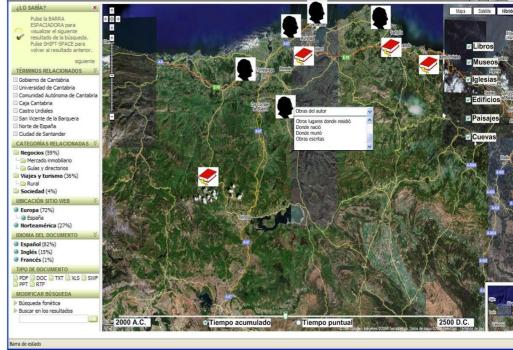
Courtesy of Anita de Waard, Elsevier, Christiaan Fluit, Aduna, and Frank van Harmelen, VU Amsterdam (SWEO Use Case)



Semantic portal for cultural heritage





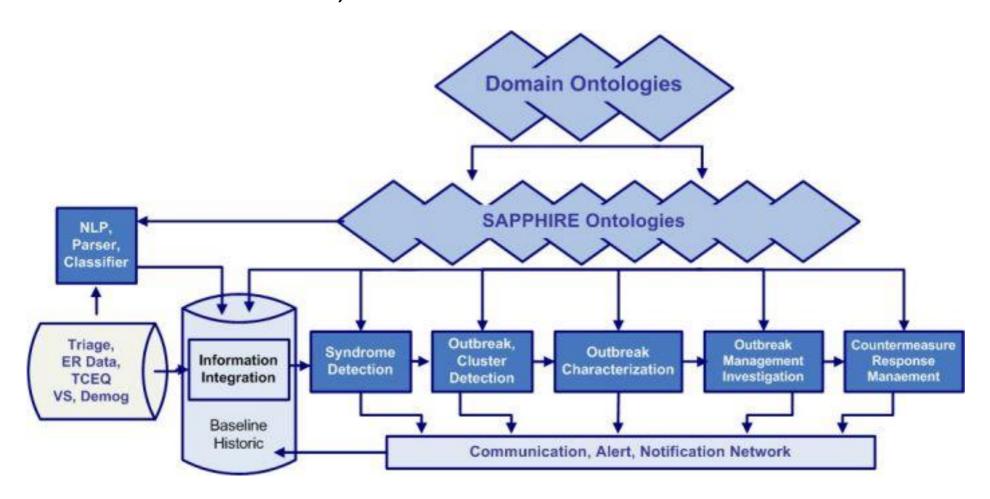


Courtesy of Francisca Hernández, Fundación Marcelino Botín, and Richard Benjamins, iSOCO, (SWEO Case Study)



Public Health Surveillance (Sapphire)

 Integrated biosurveillance system (biohazards, bioterrorism, disease control, etc)

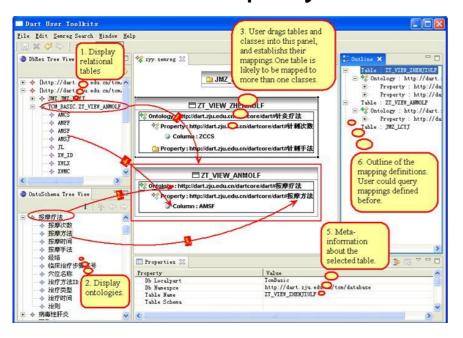


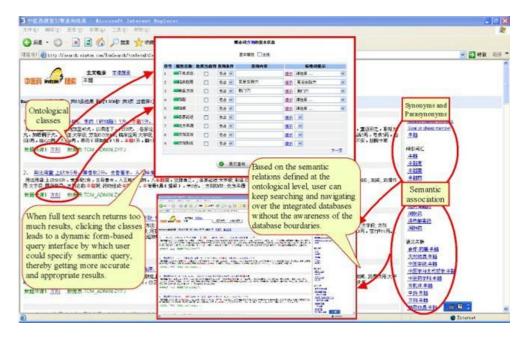
Courtesy of Parsa Mirhaji, School of Health Information Sciences, University of Texas (SWEO Case Study)



Integrate Knowledge for Chinese Medicine

- Integration of a large number of relational databases (on traditional Chinese medicine) using a Semantic Layer
 - around 80 databases, around 200,000 records each
- A visual tool to map databases to the semantic layer using a specialized ontology
- Form based query interface for end users





Courtesy of Huajun Chen, Zhejiang University, (SWEO Case Study)

