

5-2007

Semantic Web: Technologies and Applications for the Real-World

Amit P. Sheth

Wright State University - Main Campus, amit@sc.edu

Follow this and additional works at: <https://corescholar.libraries.wright.edu/knoesis>



Part of the [Bioinformatics Commons](#), [Communication Technology and New Media Commons](#), [Databases and Information Systems Commons](#), [OS and Networks Commons](#), and the [Science and Technology Studies Commons](#)

Repository Citation

Sheth, A. P. (2007). Semantic Web: Technologies and Applications for the Real-World. .
<https://corescholar.libraries.wright.edu/knoesis/640>

This Presentation is brought to you for free and open access by the The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) at CORE Scholar. It has been accepted for inclusion in Kno.e.sis Publications by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

Semantic Web: Technologies and Applications for the Real-World

Amit Sheth

LexisNexis Ohio Eminent Scholar

Kno.e.sis Center

Wright State University

<http://knoesis.wright.edu>



kno.E.sis

COLLECTING THE DOTS | CONNECTING THE DOTS

Semantic Web: Technologies and Applications for the Real-World

Susie Stephens
Principal Research Scientist

Lilly

Answers That Matter.

Tutorial Outline

Introduction to the Semantic Web	1.30-2.00pm
Real-World Applications (1): Enabling Technologies and Experiences	2.00-3.00pm
Break	3.00-3.30pm
Real-World Applications (2): Details and War Stories, Case Studies	3.30-4.30pm
Semantic Web in Practice	4.30-5.00pm

Introduction to the Semantic Web

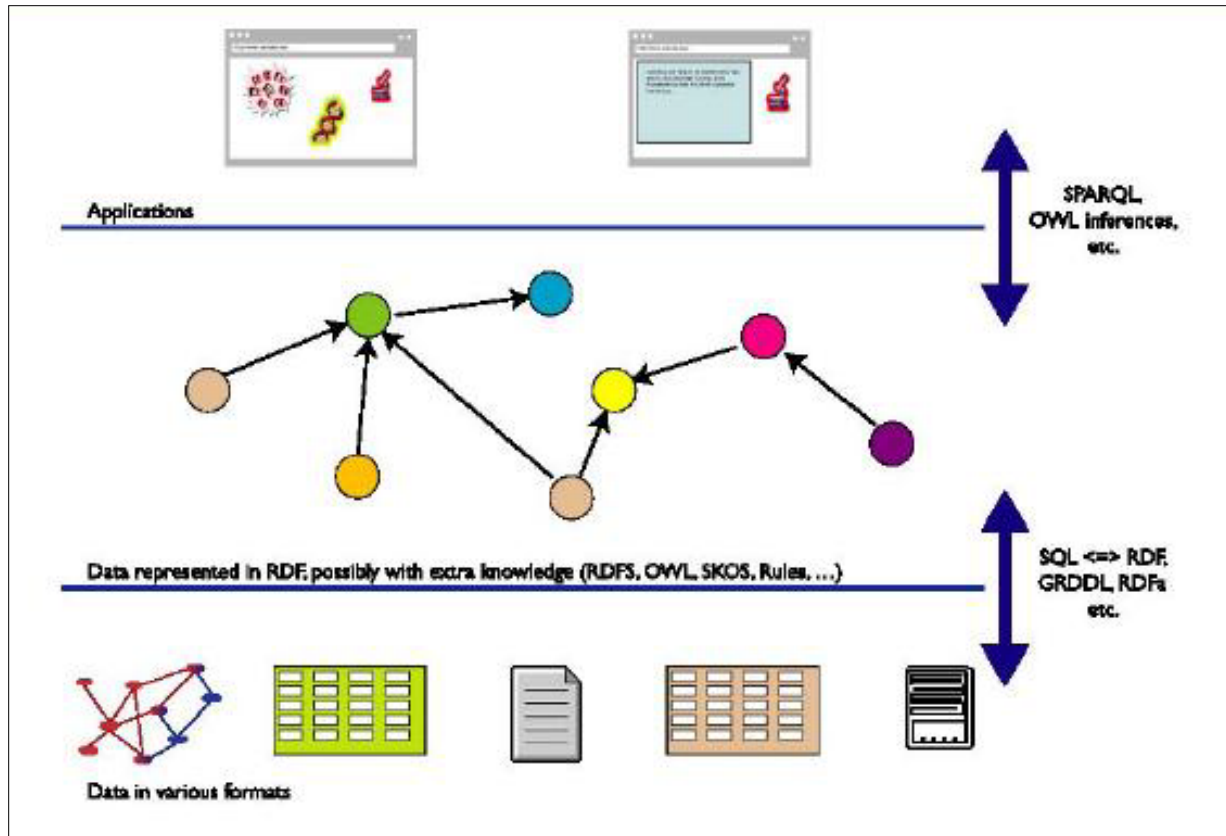
Agenda

- Characterizing the Semantic Web
- Semantic Web Standards
- Semantic Web Capabilities

Characterizing the Semantic Web

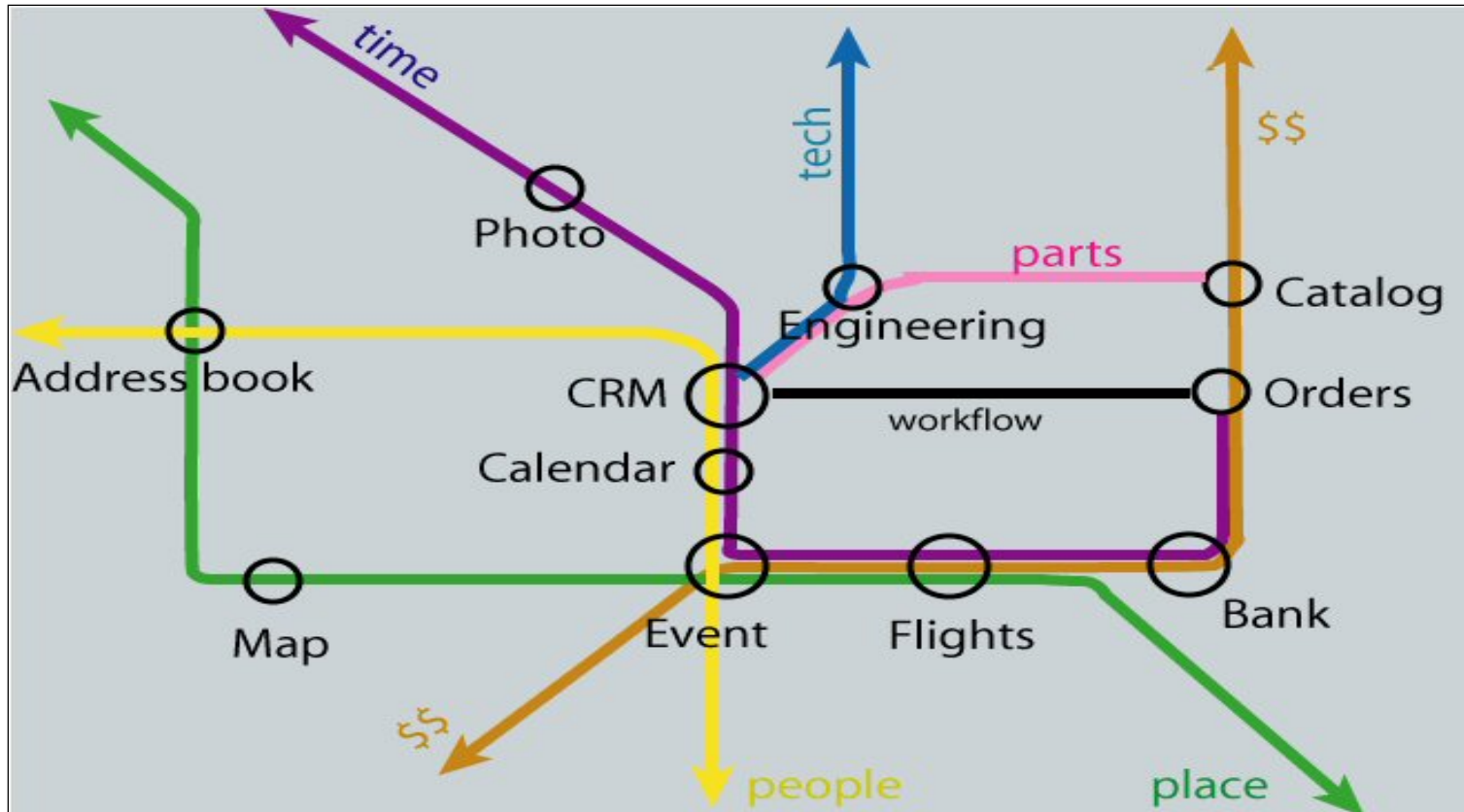
- Semantic Web is an interoperability technology
- An architecture for interconnected communities and vocabularies
- A set of interoperable standards for knowledge exchange

Creating a Web of Data



Source: Ivan Herman

Mashing Data

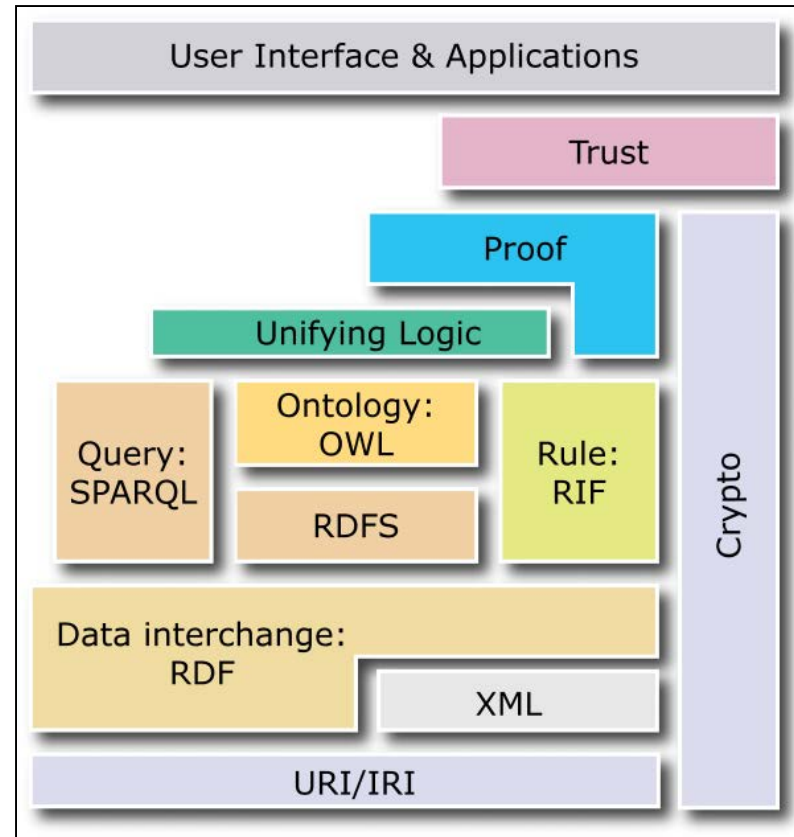


Source: W3C

Drivers for the Semantic Web

- Business models develop rapidly these days, so infrastructure that supports change is needed
- Organizations are increasingly forming and disbanding collaborations
- Data is growing so quickly that it is no longer possible for individuals to identify patterns in their heads
- Increasing recognition of the benefits of collective intelligence

Semantic Web Technologies

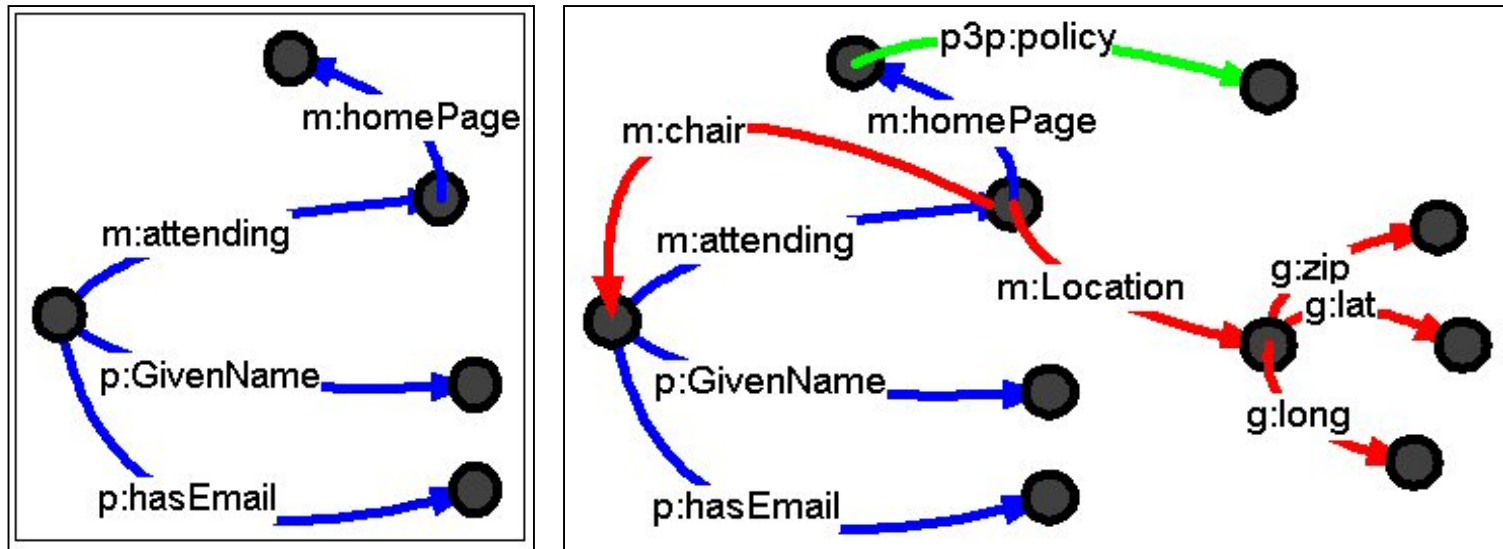


Source: W3C

Resource Description Framework (RDF)

- RDF became a W3C standard recommendation in 2004
- RDF is a language for representing information about resources in the Web
- Common framework for expressing information
- Information may be made available to applications other than those for which it was originally created
- Any information in RDF can be connected to any other information in RDF
- Resources are identified with Uniform Resource Identifiers (URIs)

Resource Description Framework (RDF)



Source: W3C

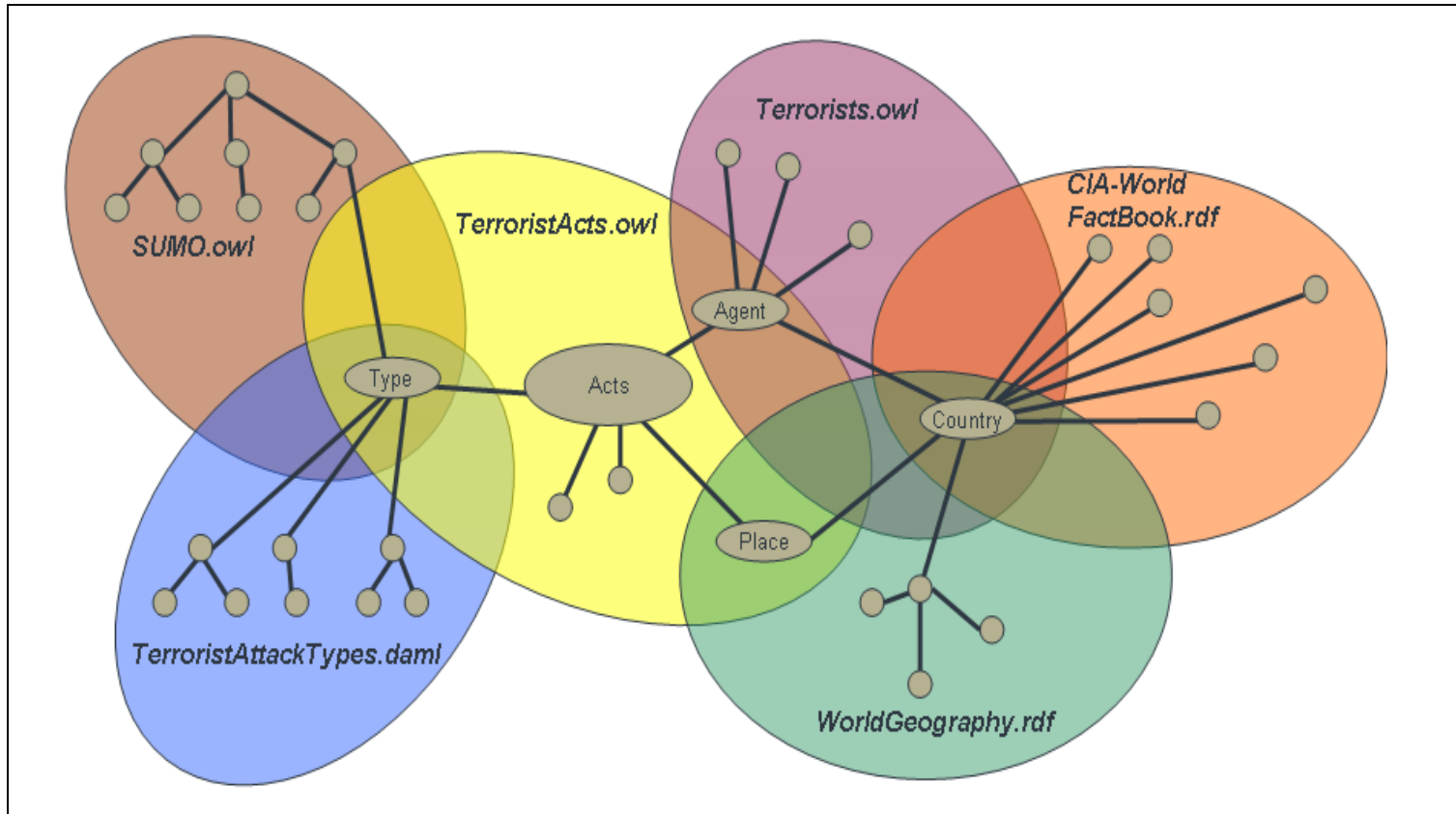
RDF Schema (RDFS)

- Vocabulary for describing properties and classes of RDF resources
- Semantics for hierarchies of properties and classes
- A resource may belong to several classes

Web Ontology Language (OWL)

- OWL became a W3C standard recommendation in 2004
- Explicitly represents meaning of terms in vocabularies and the relationships between those terms
- Separate layers have been defined balancing expressibility vs. implementability (OWL Lite, OWL DL, OWL Full)
- Supports inferencing

Merging Ontologies

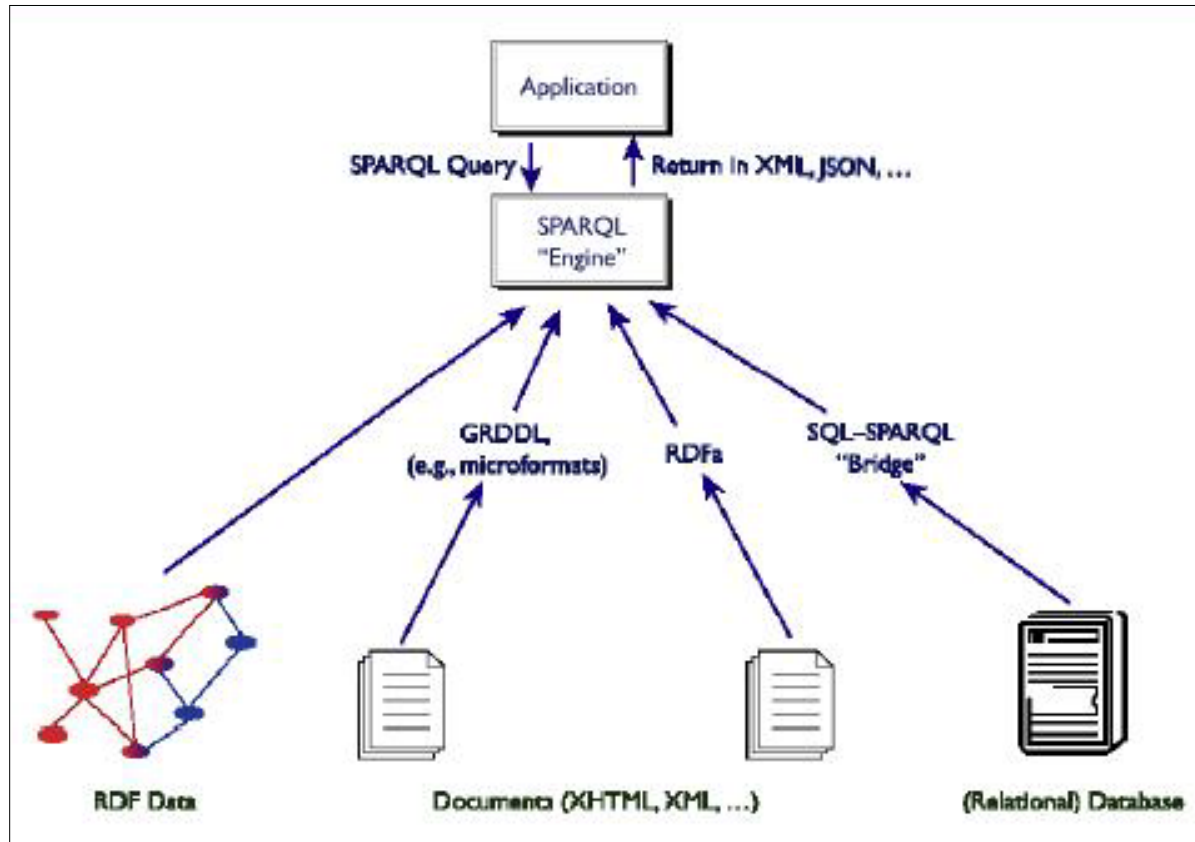


Source: Siderean Software

SPARQL

- SPARQL is expected to become a W3C standard recommendation in Q3 2007
- It is a query language based on graph patterns
- Protocol layer for using SPARQL over HTTP
- There are SPARQL endpoints on the Web
- SPARQL can be used to construct graphs

SPARQL as a Unifying Source



Source: Ivan Herman

Gleaning Resource Descriptions from Dialects of Languages (GRDDL)

- GRDDL is a mechanism for Gleaning Resource Descriptions from Dialects of Languages
- Introduces markup based on existing standards for declaring that an XML document includes data compatible with RDF and for linking to algorithms for extracting this data from the document
- GRDDL should reach W3C standard recommendation in Q2 2007

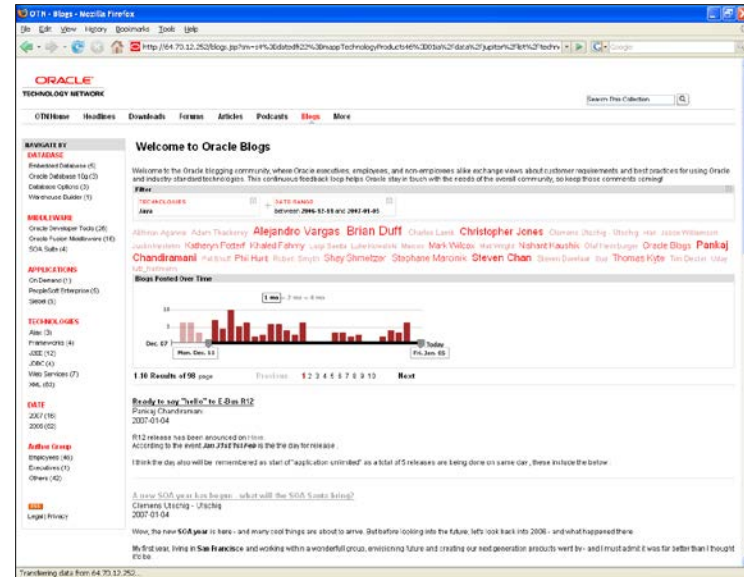
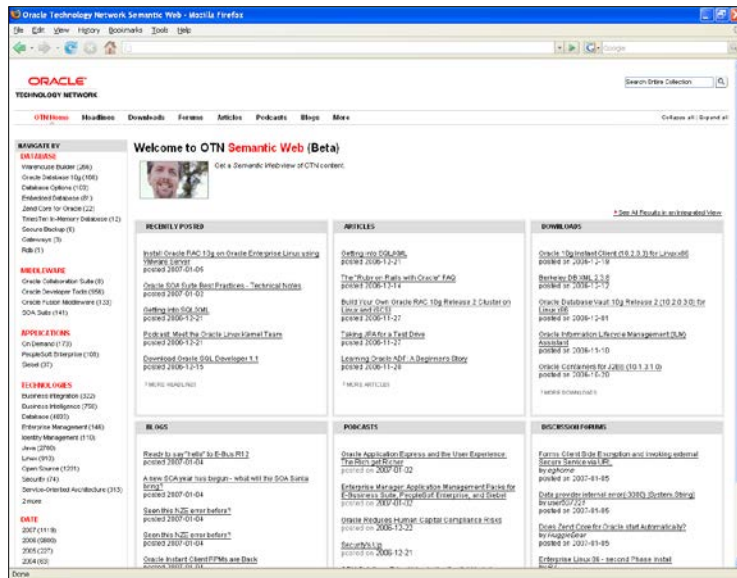
Rules

- OWL DL and OWL Lite are based on Description Logic; there are things that DL cannot express
 - Attempts to combine ontologies and rules include RuleML, SWRL, cwm, etc.
- There is also an increasing number of rule-based systems that want to interchange rules
- The W3C Rule Interchange Format Working Group is focused on this area

Semantic Annotations for WSDL (SAWSDL)

- Develop a mechanism to enable semantic annotation of Web service descriptions
- Using the WSDL 2.0 extension mechanism to build simple and generic support for adding service descriptions for Web services
- This W3C standard is at Candidate Recommendation
- Tools and Use cases available (Google: SAWSDL)

Enhanced Enterprise Search




Source: Oracle OTN

Improved Reliability of Search Results



Source: Segala

Web of Data




"Information wants to be free."
- Stewart Brand, 1984

Introducing Freebase.
Freebase is an open, shared database of the world's knowledge.

"Need for Speed" by Amnemonia flickr

Free + Database = Freebase
It's about film, sports, politics, music, science and everything else all connected together. Our contributors are collecting data from all over the internet to build a massive, collaboratively-edited database of cross-linked data. It's a big job and we're just getting started.

Share, reuse, remix.
We want to make it possible for you to add high quality structured information to your websites, mashups and applications without worrying about restrictive licenses. All data is licensed Creative Commons Attribution. We only ask that you link back to us.



Source: Freebase

Data Mashups

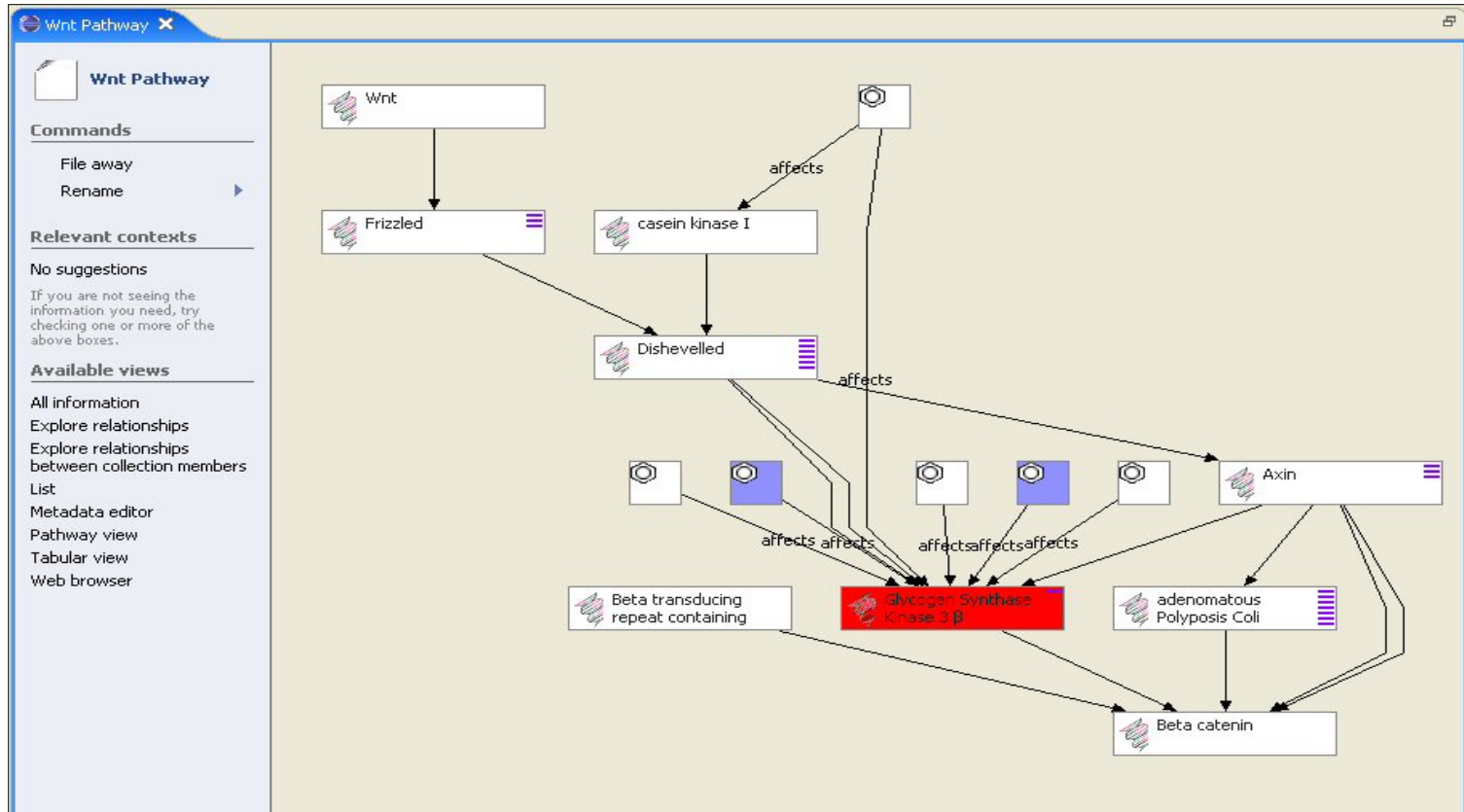
The screenshot displays the Haystack Eclipse Platform interface with two main views:

- GSK3beta Topic:** This view features a 'Target overview' diagram where several entities (ce-DBAP Lead, OSB..., OA..., CHI..., AKAPsullo NCE) are connected to a central 'GSK-3beta' node via arrows labeled 'targets'. Below this diagram are sections for 'Group members', 'Alternative diseases', 'Primary disease', 'libraries used', 'clinical studies', and 'relevant items'.
- Wnt Pathway:** This view shows a hierarchical pathway diagram. At the top is 'Wnt', which leads to 'Frizzled' and 'casein kinase I'. Both of these lead to 'Dishevelled'. From 'Dishevelled', the pathway branches into 'Axin', 'Glycogen Synthase Kinase 3 β', and 'Beta catenin'. 'Axin' and 'Glycogen Synthase Kinase 3 β' both lead to 'Beta catenin'. 'Beta catenin' then leads to 'Beta transducing repeat containing'.

The interface includes a menu bar (File, Edit, Navigate, Search, Project, Haystack, Run, Window, Help) and a toolbar with various navigation icons. The left sidebar contains 'Commands' and 'Relevant contexts' sections.

Source: BioDASH

Data Mashups



Source: BioDASH

Document Management

The screenshot displays the MySWAN web application interface. At the top, the logo "MySWAN" is followed by the tagline "semantic web applications in neuromedicine" and "Brought to you by Alzforum and MIND Informatix". Navigation links include "Log In", "Register", "My Profile", and "About". A secondary navigation bar contains "Home", "Swan Data", and "Help".

The main content area is titled "Hypothesis" and features the following fields:

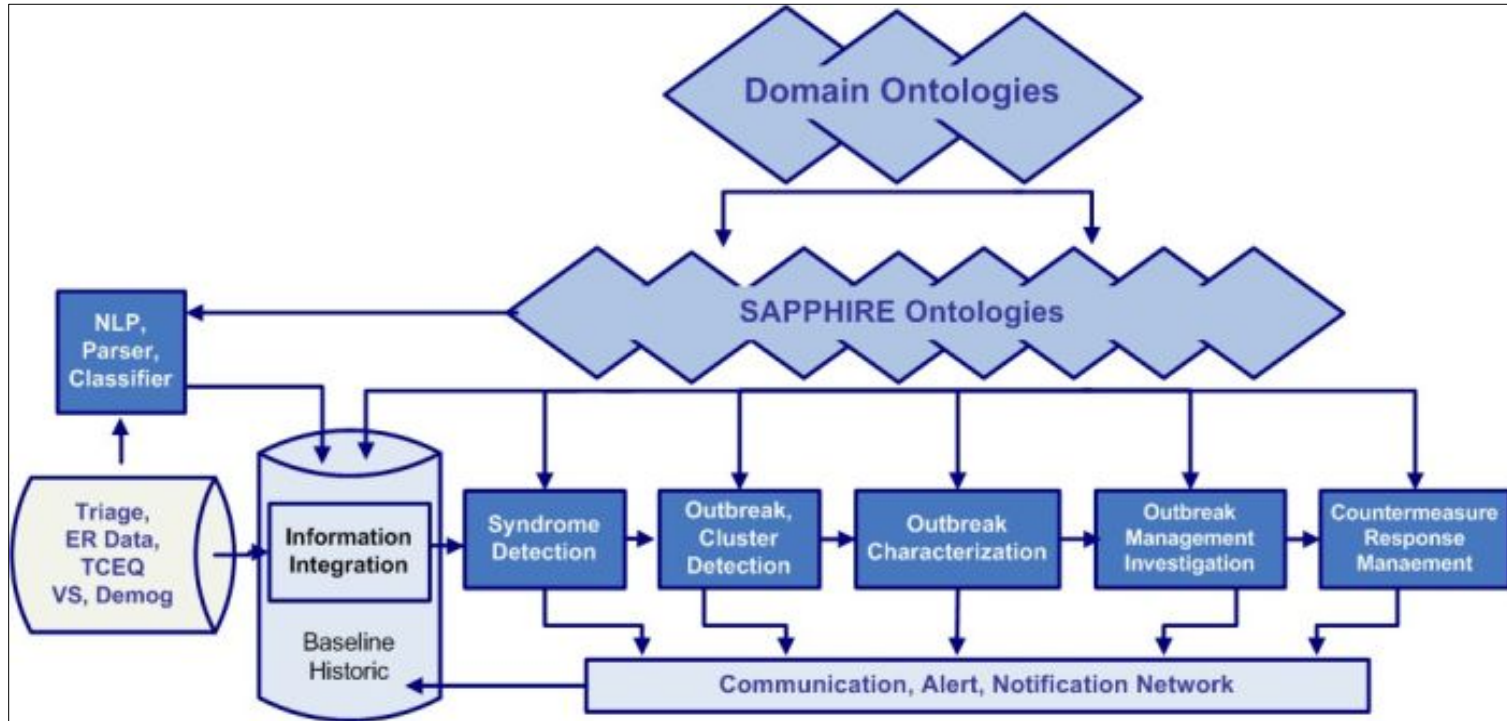
- Title:** An alternative interpretation of the amyloid Abeta hypothesis with rega
- Abstract:** Alzheimer's disease is a complex neurodegenerative process that is believed to be due to the accumulation of short, hydrophobic peptides derived from amyloid precursor proteins by proteolytic cleavage. It is widely believed that these Abeta peptides are secreted
- Authors:** Last Name First Name Middle Name
- Comments:** Full-text summary and discussion located at <http://www.alzforum.org/res>
- Claims:** A section with a "Drop Claims Here" button and a list of 16 numbered claims, each followed by a small icon.

The left sidebar contains a "Search" box and a "SWAN Data" menu with the following items: Collections, Publications, Manuscripts, Websources, Datasets, Hypotheses (with a sub-item "An alternative interpretation of th"), Claims, and Categories (with sub-items: Behavioral Evidence, EvidenceClass, Biomarker Evidence, Drug Target Evidence, Genetic Evidence, Initial Conditions, Pathophysiologic Evidence, Convergence, Progression, Animal Model Evidence, Pathway Evidence, Downstream Events, Perturbation, Epidemiologic Evidence, ClinicalTrial Evidence, Etiologic Evidence). A "Collaboration" button is at the bottom of the sidebar.

The right sidebar contains a "Toolbox" with the following options: Create, Import, Search, and Export data.

Source: MGH

Semantic Data Warehouse



Source: University of Texas

Heterogeneous Data Aggregation

- Ordnance Survey maintains the definitive mapping data of Great Britain
- Use ontologies to bridge semantically diverse sources of data, e.g., Hydrology, Administrative Geography, Buildings and Places, etc.
- Efficient queries via the ontology or directly in RDF
- Advantages include efficient data integration, data repurposing, better quality control and classification

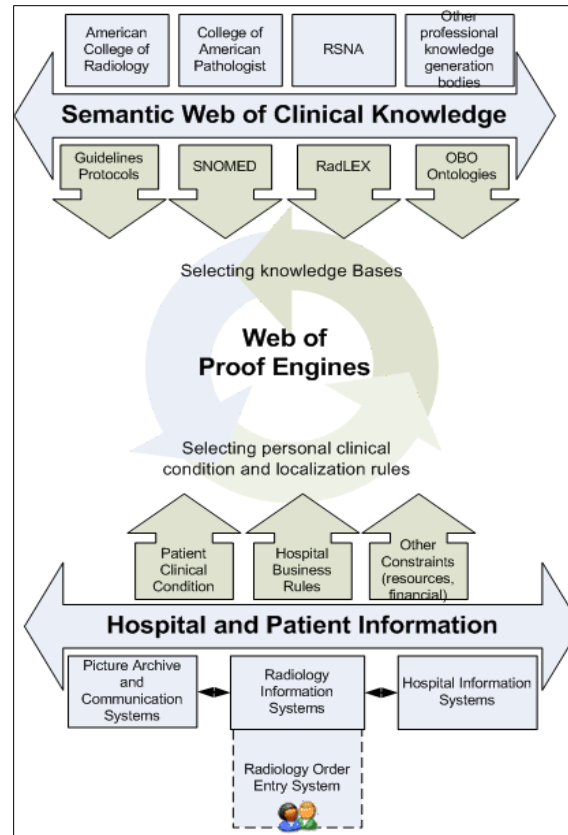
Integration with Semantic Mediation

- BT uses semantic descriptions of system interfaces and messages to support integration of Operational Support Systems (OSS)
- Internet Service Providers integrate their OSS with those of BT (via a gateway)
- The approach helps overcome the increasing complexity of supply chains, reduces costs and time-to-market, ontologies allow for a reuse of services

Natural Language Query of Applications

- Tata Consulting Services provides a natural language interface to business applications
- Users pose questions and invoke actions using natural language
- An OWL ontology aids in the retrieval of relevant data and concepts
- Advantages include distinct semantics for various domain concepts, external concepts linked in, simple interface

Decision Support



Source: AGFA

Summary

- The Semantic Web provides functionality of interest to business, scientific and Web communities
- Capabilities are provided in data integration, search, semantic mediation, decision support, etc.
- Semantic Web standards are maturing
- The Semantic Web is increasingly being adopted

Real World Applications

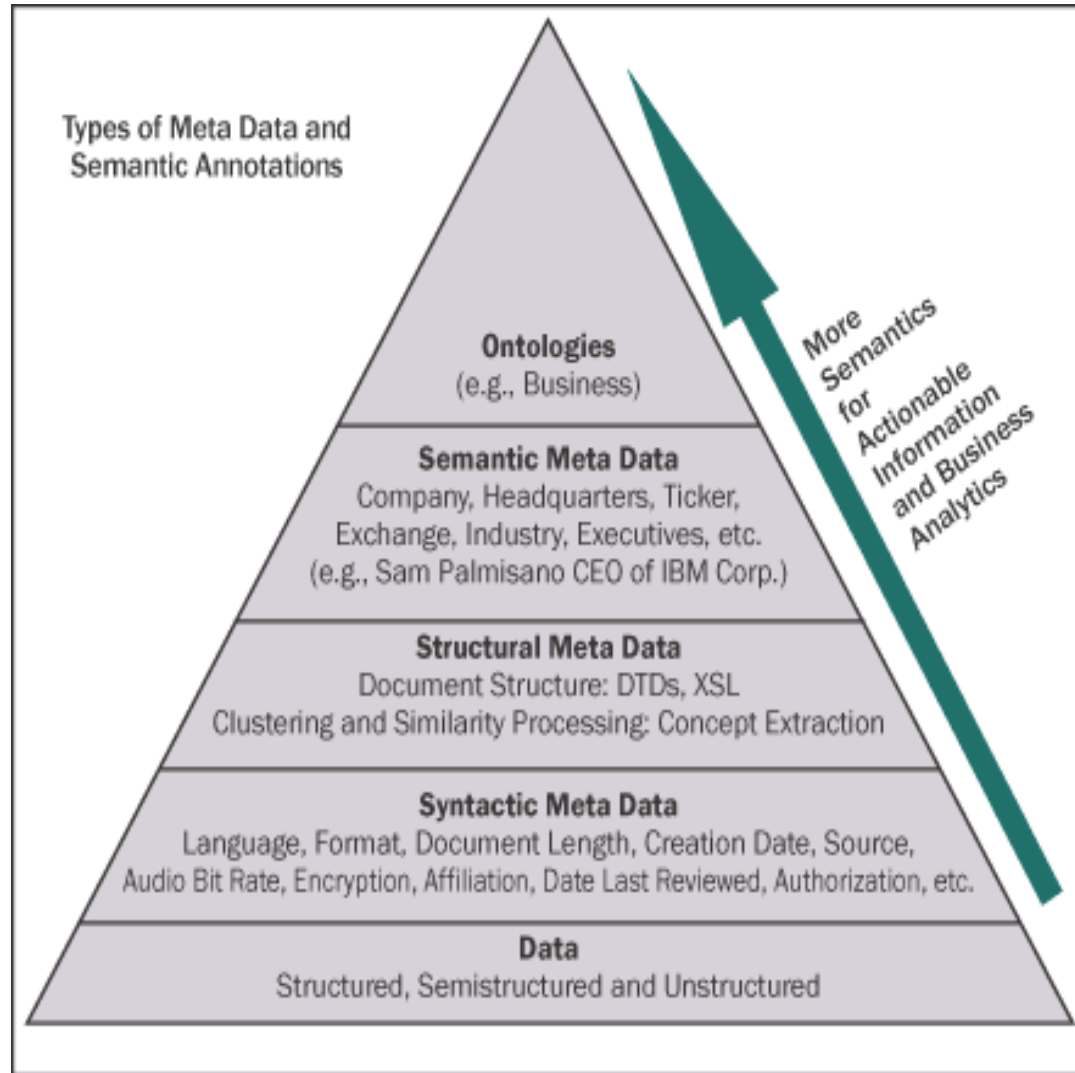
with Enabling Capabilities/Technologies

Semantic Web Application Lifecycle

Enablers and Techniques

Syntax, Structure and Semantics

Semantics:
**Meaning &
Use of Data**



Describing Semantic Web to Nontechnical Users

Labeling data on Web so that both humans and machines can more effectively use them

Associating meaning to data that machines can understand so as to achieve lot more automation and off-load more work to machines

Exploiting common vocabulary and richer modeling of subject area for much better integration of data

Enablers and Techniques

Ontology: Agreement with Common Vocabulary & Domain Knowledge; Schema + Knowledge base

Semantic Annotation (metadata Extraction): Manual, Semi-automatic (automatic with human verification), Automatic

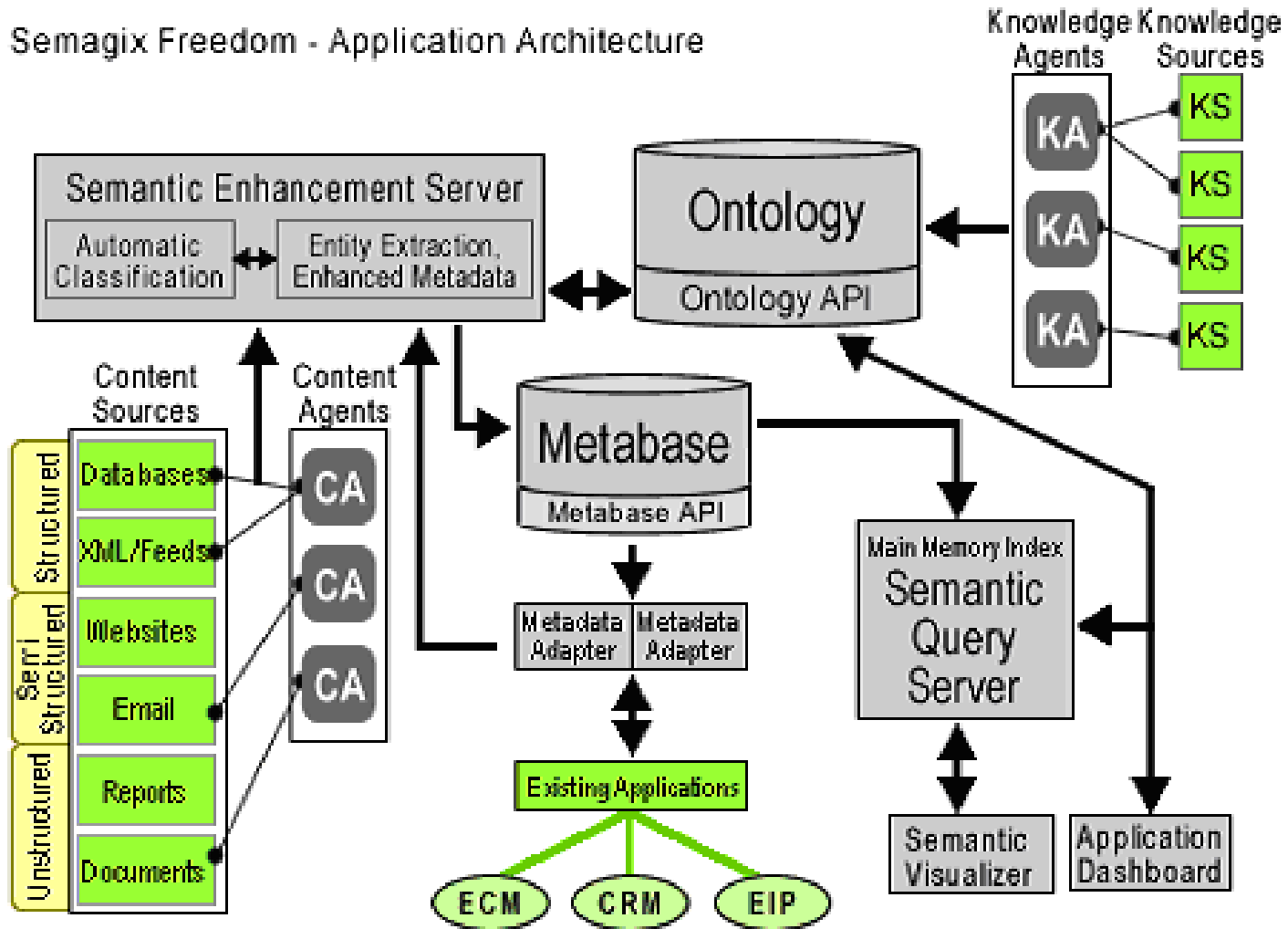
Reasoning/computation: semantics enabled search, integration, complex queries, analysis (paths, subgraph), pattern finding, mining, hypothesis validation, discovery, visualization

A Typical Enterprise SW Application Lifecycle

- Build Ontology
 - **Build Schema (model level representation)**
 - **Populate with Knowledgebase (people, location, organizations, events)**
- Automatic Semantic Annotation (Extract Semantic Metadata)
 - **Any type of document, multiple sources of documents**
 - **Metadata can be stored with or sparely from documents**
- Applications: search (ranked list of documents of interest (semantic search), integrate/portal, summarize/explain, analyze, make decisions
 - **Reasoning techniques: graph analysis, inferencing**

Types of content/documents, Use of standards, Scalability, Performance

Semagix Freedom Architecture: for building ontology-driven information system



Managing Semantic Content on the Web: <http://portal.acm.org/citation.cfm?id=613729>

Building ontology

Three broad approaches:

Option 1: social process/manual: many years, committees

- Can be based on metadata standard

**Option 2: automatic taxonomy generation (statistical clustering/NLP):
limitation/problems on quality, dependence on corpus, naming**

**Option 3: Descriptive component (schema) designed by domain experts;
Description base (assertional component, extension) by automated
processes**

Option 2 is being investigated in several research projects;

Option 3 is currently supported by technologies such as Semagix Freedom

Ontology Examples

Time, Space

Gene Ontology, Glycomics

Pharma Drug, Treatment-Diagnosis

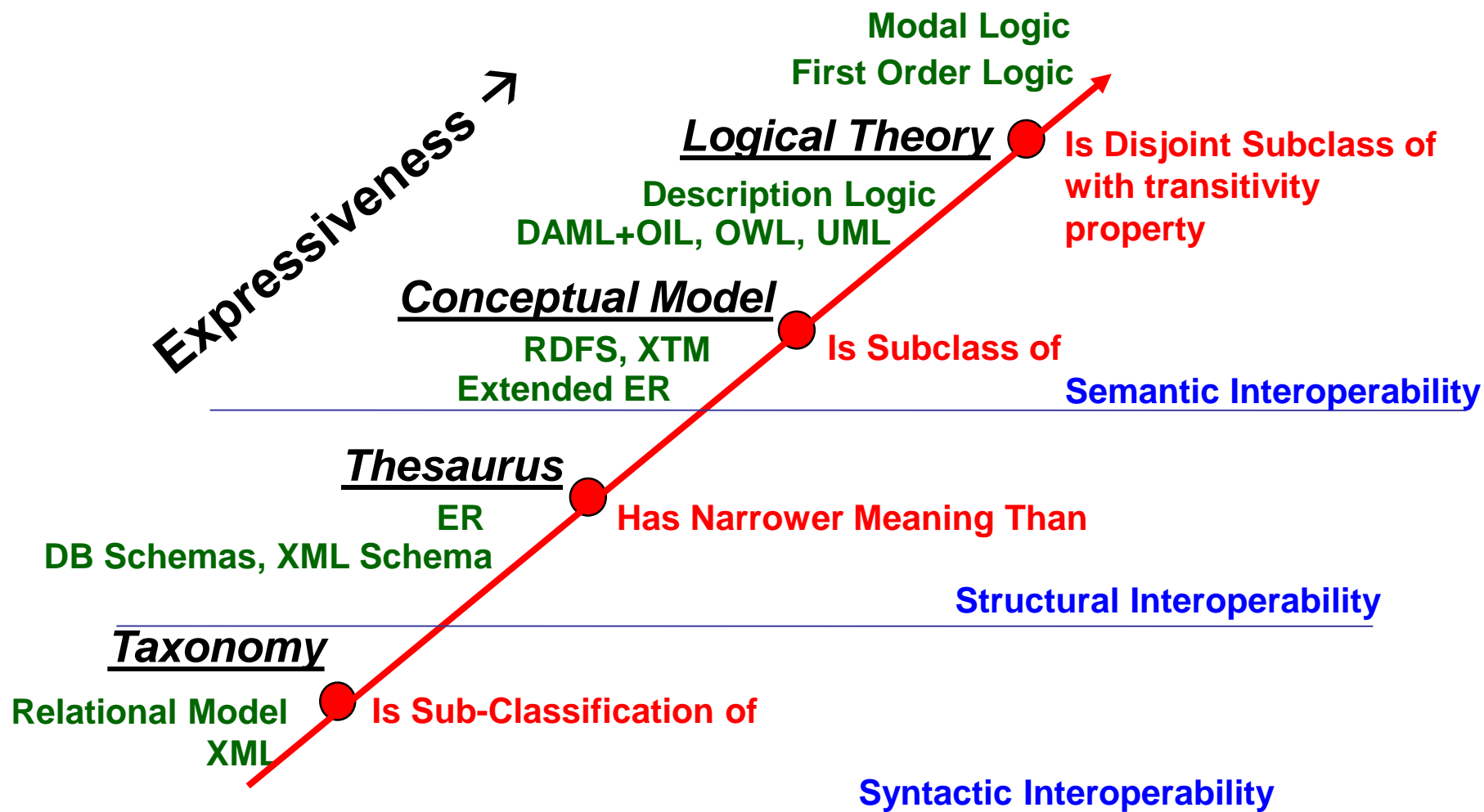
Repertoire Management

Equity Markets

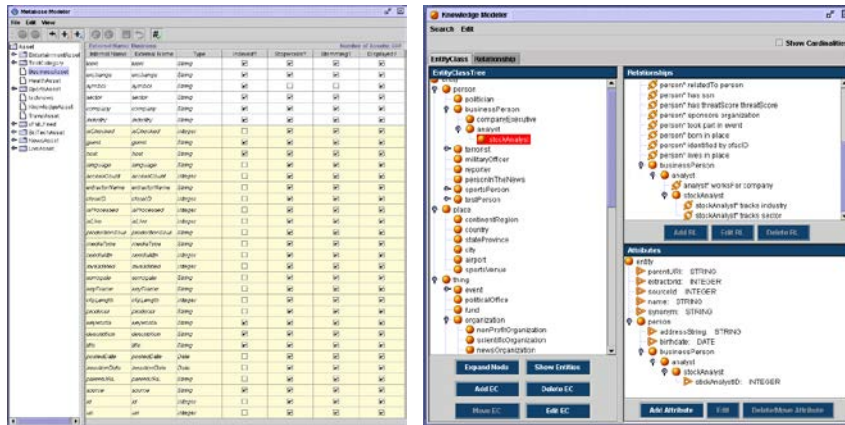
Anti-money Laundering, Financial Risk, Terrorism

Can be Public, Government, Limited Availability, Commercial

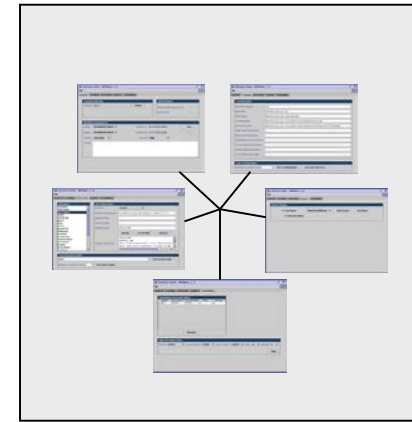
Ontology Language/ Representation Spectrum



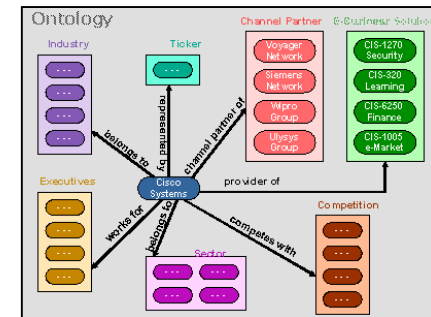
Ontology Creation and Maintenance Steps (Approach 1)



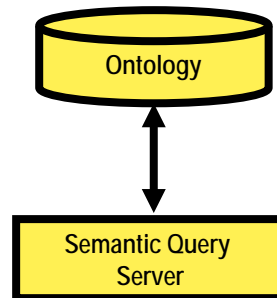
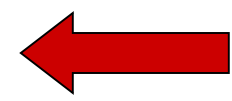
1. Ontology Model Creation (Description)



2. Knowledge Agent Creation



3. Automatic aggregation of Knowledge



4. Querying the Ontology



© Semagix, Inc.

Some observations on Ontology Development and Maintenance

- Type of domain
Modeling of human centric world (sports, entertainment, legal, financial services)
Versus
Natural world (life sciences, astronomy, ...)
- Schema
- Sources of knowledge

GlycO

*is a focused ontology for the description of glycomics
models the biosynthesis, metabolism, and biological
relevance of complex glycans*

models complex carbohydrates as sets of simpler
structures that are connected with rich relationships

An ontology for structure and function of Glycopeptides

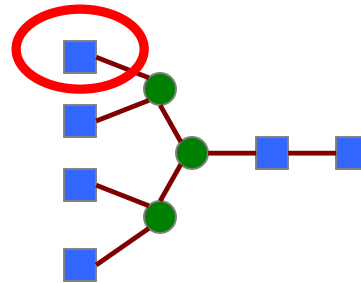
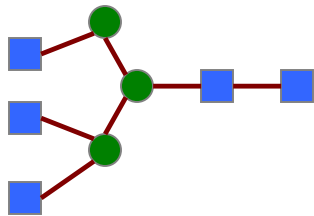
Published through the National Center for Biomedical
Ontology (NCBO) and Open Biomedical Ontologies (OBO)

See: [GlycoDoc](#), [GlycO](#)

GlycO

Challenge – model hundreds of thousands of complex carbohydrate entities

But, the differences between the entities are small (E.g. just one component)



How to model all the concepts but preclude redundancy
→ ensure **maintainability**, **scalability**



GlycO population

Assumption: with a large body of background knowledge, learning and extraction techniques can be used to assert facts.

Asserted facts are compositions of individual building blocks

Because the building blocks are richly described, the extracted larger structures will be of high quality

GlycO Population

Multiple data sources used in populating the ontology

- o KEGG - Kyoto Encyclopedia of Genes and Genomes
- o SWEETDB
- o CARBBANK Database

Each data source has a different schema for storing data

There is significant overlap of instances in the data sources

Hence, entity disambiguation and a common representational format are needed

Diverse Data From Multiple Sources Assures Quality

Democratic principle

Some sources can be wrong, but not all will be

More likely to have homogeneity in correct data than in erroneous data

Ontology population workflow

Semagix Freedom knowledge
extractor

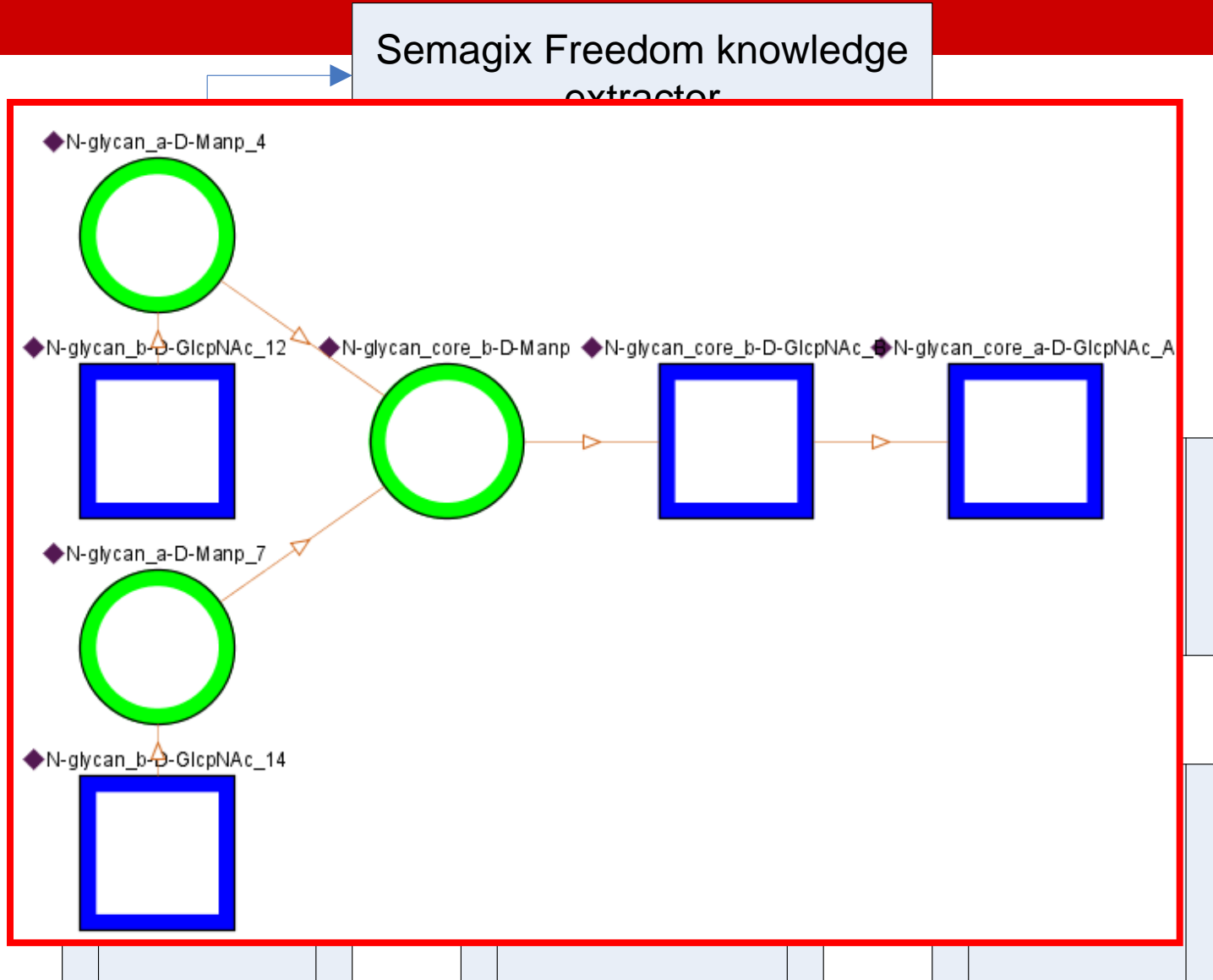
```
<Glycan>
  <aglycon name="Asn"/>
  <residue link="4" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="GlcNAc">
    <residue link="4" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="GlcNAc">
      <residue link="4" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="Man" >
        <residue link="3" anomeric_carbon="1" anomer="a" chirality="D" monosaccharide="Man" >
          <residue link="2" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="GlcNAc" >
            </residue>
          <residue link="4" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="GlcNAc" >
            </residue>
          </residue>
        <residue link="6" anomeric_carbon="1" anomer="a" chirality="D" monosaccharide="Man" >
          <residue link="2" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="GlcNAc">
            </residue>
          </residue>
        </residue>
      </residue>
    </residue>
  </residue>
</Glycan>
```

insert into
KB

Knowledge
Base

LINKS TO
GLYDE

Ontology population workflow



Diverse Data From Multiple Sources Assures Quality

Holds only, when the data in each source is independent

In the case of GlycO, the sources that were meant to assure quality were not diverse.

One original source (CarbBank) was copied by several Databases without curation

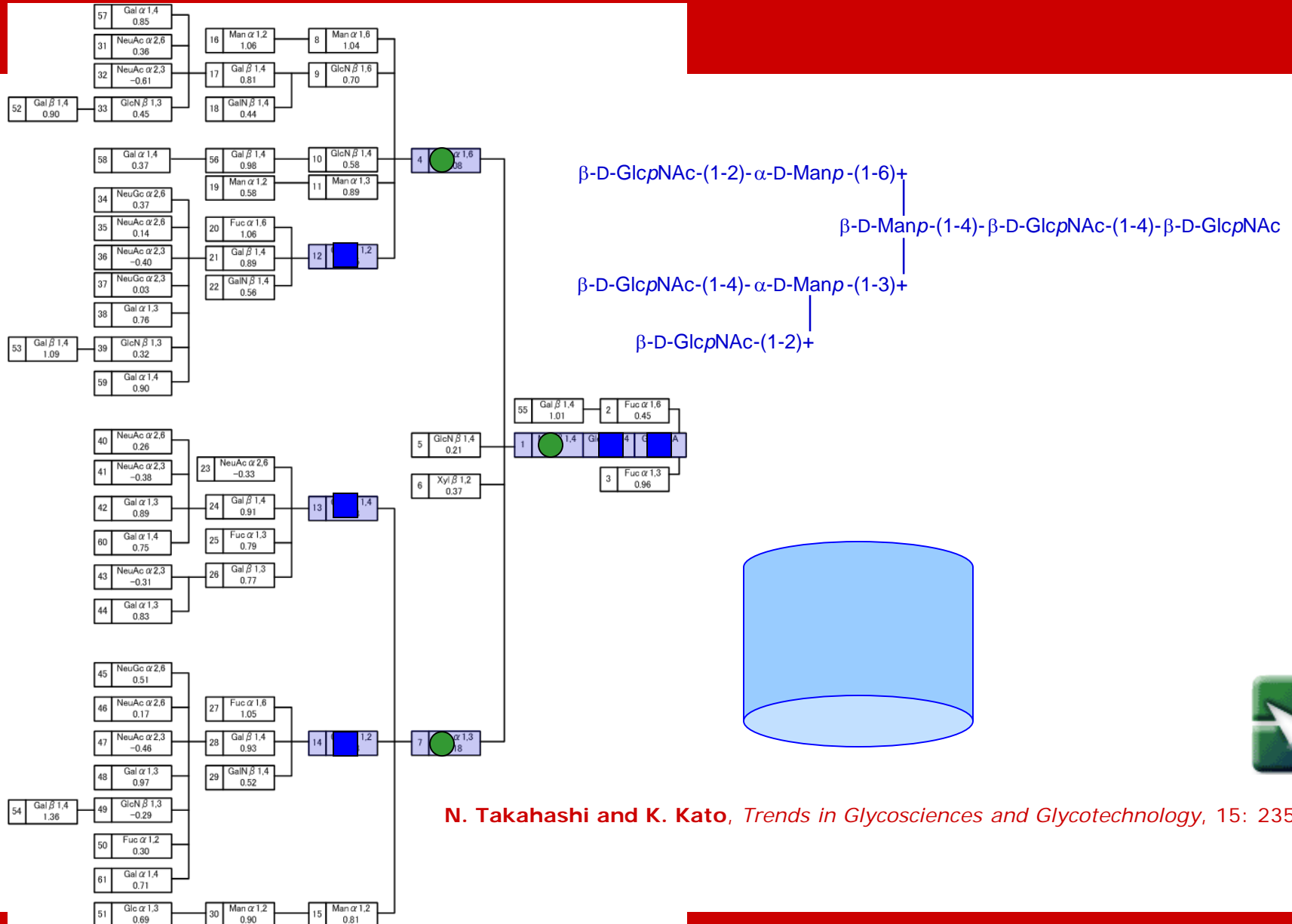
Errors in the original propagated

Errors in KEGG and CarbBank are the same

Cannot use these sources for comparison

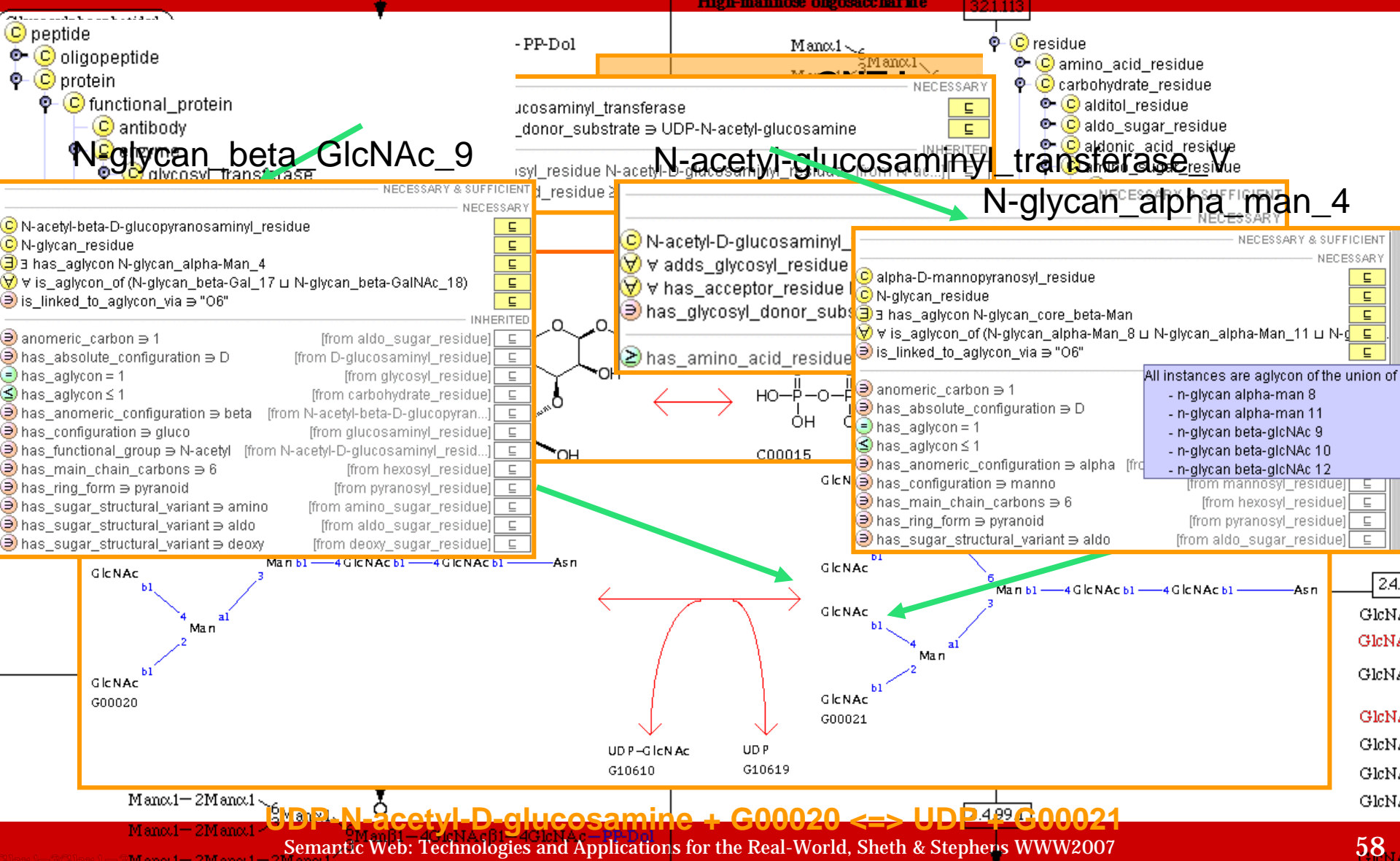
Needs curation by the expert community

GlycoTree



N. Takahashi and K. Kato, *Trends in Glycosciences and Glycotechnology*, 15: 235-251

N-Glycosylation metabolic pathway



UDP-N-acetyl-D-glucosamine + G00020 \rightleftharpoons UDP-G00021

Semantic Web: Technologies and Applications for the Real-World, Sheth & Stephens WWW2007

Pathway Steps - Glycan

The diagram illustrates the synthesis of a glycan moiety through two reaction steps:

- Reaction 1:** `reaction_ec_3.2.1_CR00001` (top) has a reactant (represented by an empty box) and a product (the glycan moiety).
- Reaction 2:** `reaction_ec_3.2.1_CR00003` (bottom) has a reactant (the glycan moiety) and a product (represented by an empty box).

The central **glycan_moiety** consists of a branched structure of green circles (sugars) and blue squares (sugars). The rightmost part of the glycan is a **peptide_moiety** (white circle) attached via a $\beta 4$ linkage.

Abundance of this glycan in three experiments:

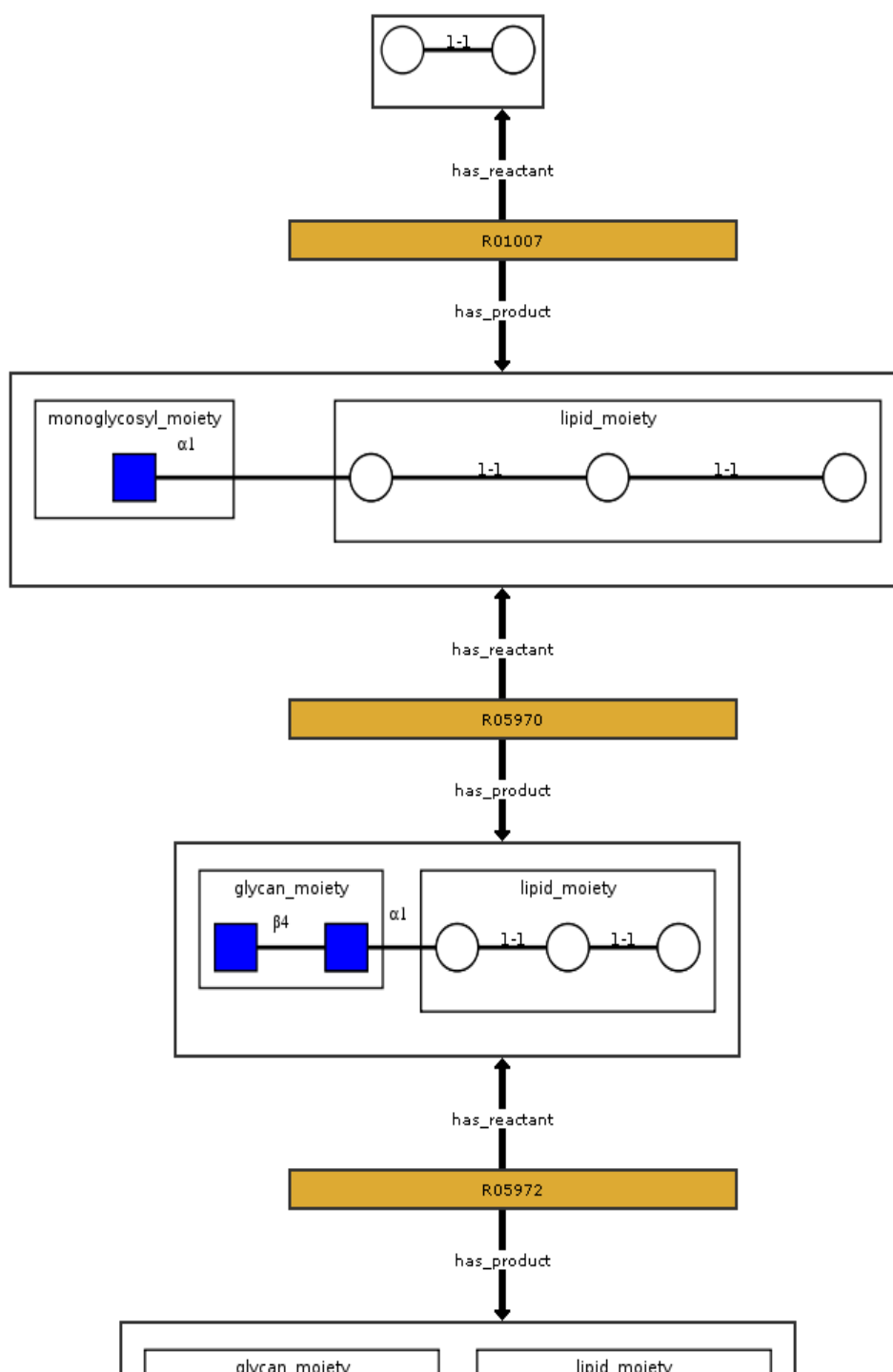
Experiment	Abundance
ES	14.95
EB	7.64
RA	11.05

Metadata for glyco peptide G10695:

- has_peptide_moiety:** representative asparagine-containing peptide moiety
- has_carbohydrate_moiety:** N-Glycan-G10695
- has_link:** N-glycan precursor moiety G10695 to representative peptide moiety link
- label:** G10695
- has_moiety:** representative asparagine-containing peptide moiety
- has_non-carbohydrate_moiety:** representative asparagine-containing peptide moiety
- has_part:** representative asparagine-containing peptide moiety
- has_part:** N-Glycan-G10695
- has_moiety:** N-Glycan-G10695

Downstream

Abundance of this glycan in three experiments



ProPreO ontology

An ontology for capturing process and lifecycle information related to proteomic experiments

Two aspects of glycoproteomics:

What is it? → identification

How much of it is there? → quantification

Heterogeneity in data generation process, instrumental parameters, formats

Need data and process provenance → ontology-mediated provenance

Hence, ProPreO models both the glycoproteomics experimental process and attendant data

Published through the National Center for Biomedical Ontology (NCBO) and Open Biomedical Ontologies (OBO)

More at: <http://knoesis.wright.edu/research/bioinformatics/>

Semantic Annotation

Shallow Annotation: SemTag

Intermediate Annotation:

- **SEE/Semagix**: <http://lsdis.cs.uga.edu/library/download/HSK02-SEE.pdf>
- **Ontotext**: <http://www.ontotext.com/kim/semanticannotation.html>

Advanced: Schema-driven Relationship Extraction Engine (**SCREEN**): <http://knoesis.wright.edu/research/discovery>

Extracting a Text Document: Syntactic approach

INCIDENT MANAGEMENT SITUATION REPORT

LAYOUT

Friday August 1, 1997 - 0530 MDT

NATIONAL PREPAREDNESS LEVEL II

CURRENT SITUATION: Alaska continues to experience large fire activity. Additional fires have been staffed for structure protection.

SIMELS, Galena District, BLM. This fire is on the east side of the Innoko Flats, between Galena and McGr. The fire is active on the southern perimeter, which is primarily spruce. The fire has increased in size, but was not made. The fire is 35% contained, while protection of the M.

Date => day month int ', ' int

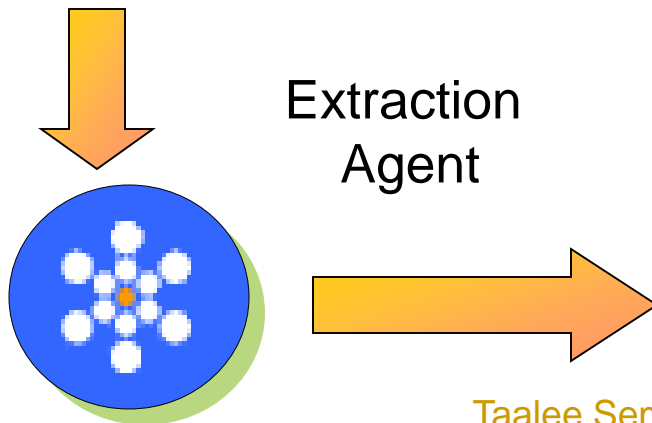
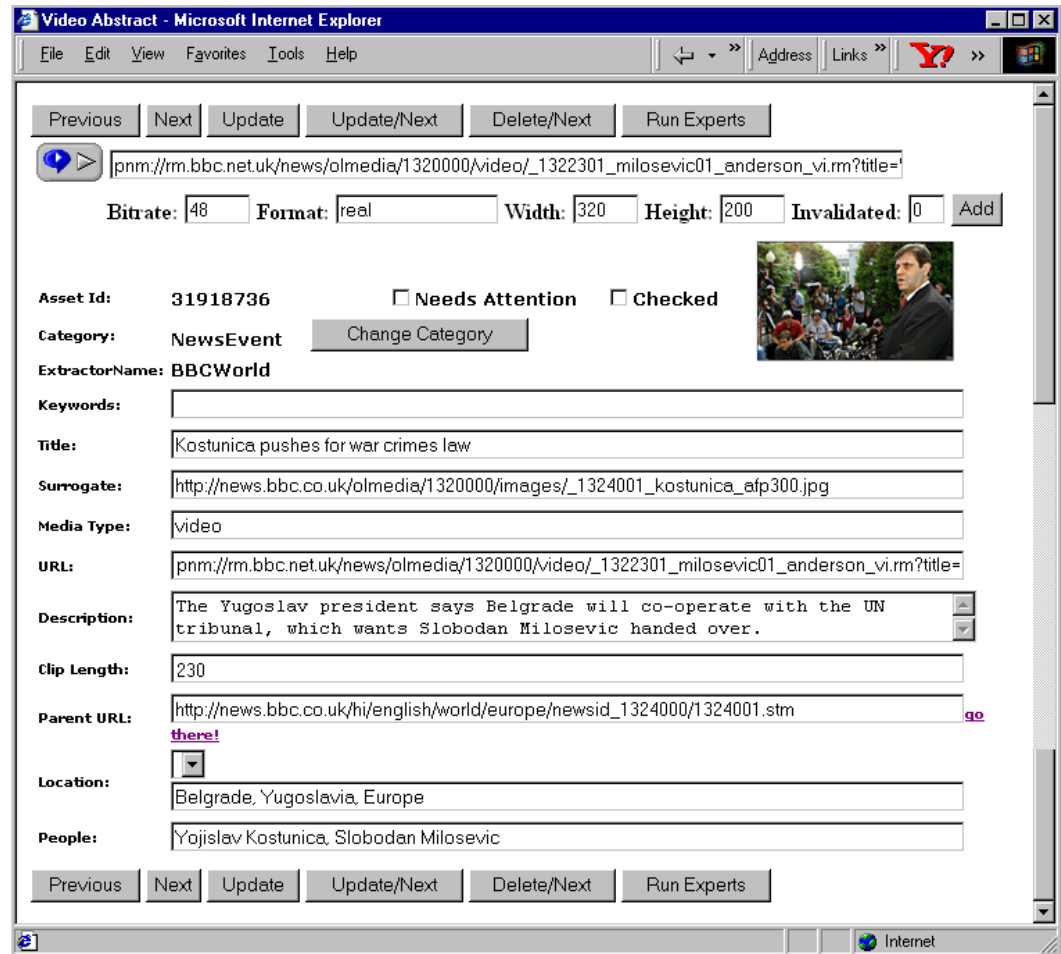
CHINIKLIK MOUNTAIN, Galena District, BLM. A Type II Incident Management Team (Wehking) is assigned to the Chiniklik fire. The fire is contained. Major areas of heat have been mopped up. The fire is contained. Major areas of heat have been mopped-up. All crews and overhead will mop-up where the fire burned beyond the meadows. No flare-ups occurred today. Demobilization is planned for this weekend, depending on the results of infrared scanning.

Taalee Extraction and Knowledgebase Enhancement

Web Page

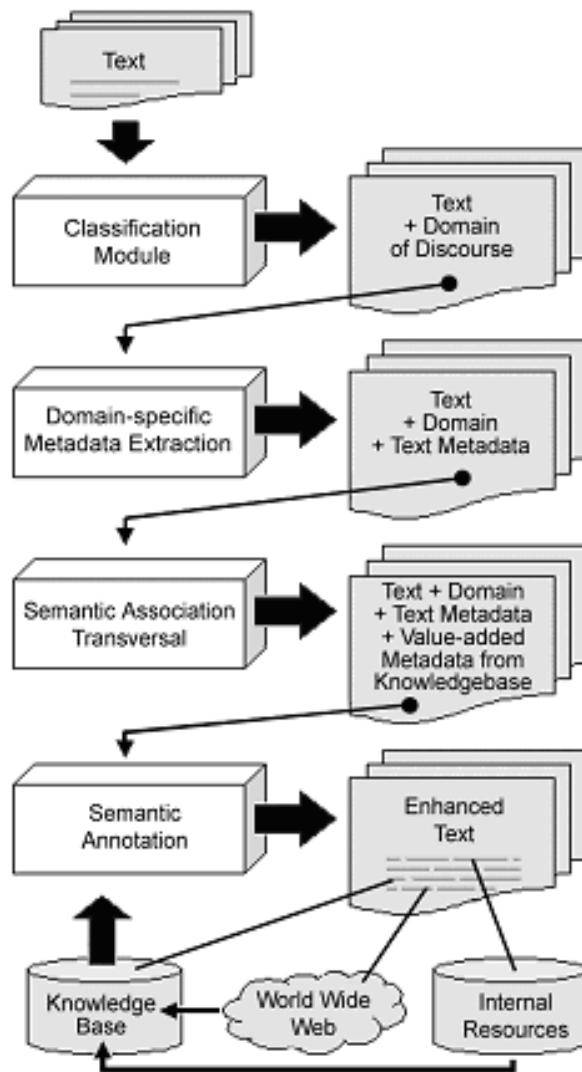


Enhanced Metadata Asset



Taalee Semantic Engine, also IBM Semantics Tools

Metadata Extraction and Semantic Enhancement



[Semantic Enhancement Engine](#)
[Hammond, Sheth, Kochut 2002]
Also, [WebFountain](#),
[KIM from OntoText](#)

Semantic Annotation - document

Blue-chip bonanza continues

company company company
Dow above 9,000 as [HP](#), [Home Depot](#) lead advance; [Microsoft](#) upgrade helps techs.

date time
August 22, 2002: 11:44 AM EDT

phrase phrase
By Alexandra Twin, CNN/Money Staff Writer

city company
[New York](#) (CNN/Money) - An upgrade of software leader [Microsoft](#) and strength in blue chips including [Hewlett-Packard](#) and [Home Depot](#) were among the factors pushing stocks higher at midday [Thursday](#), with the [Dow Jones industrial average](#) spending time above the 9,000 level.

time financial index
Around 11:40 a.m. ET, the [Dow Jones industrial average](#) gained 65.06 to 9,022.09, continuing a more than 1,300-point resurgence since July 23. The [Nasdaq](#) composite gained 9.12 to 1,418.37.

financial index
[The Standard & Poor's 500 index](#) rose 9.61 to 958.97.

company stockSym \$ \$
[Hewlett-Packard](#) ([HPQ](#): up \$0.33 to \$15.03, Research, Estimates) said a report shows its share of the printer market grew in the second quarter, although another report showed that its share of the computer server market declined in [Europe](#), the [Middle East](#) and [Africa](#).

company stockSym \$ \$
[Home Depot](#) ([HD](#): up \$1.07 to \$33.75, Research, Estimates) was up for the third straight day after topping fiscal second-quarter earnings estimates on Tuesday.

tech category company
Tech stocks managed a turnaround. [Software](#) continued to rise after [Salomon Smith Barney](#) upgraded No. 1 software maker [Microsoft](#) ([MSFT](#): up \$0.55 to \$52.83, Research, Estimates) to "outperform" from "neutral" and raised its price target to \$59 from \$56. Business software makers [Oracle](#) ([ORCL](#): up \$0.18 to \$10.94, Research, Estimates), [PeopleSoft](#) ([PSFT](#): up \$1.17 to \$20.67, Research, Estimates) and [BEA Systems](#) ([BEAS](#): up \$0.28 to \$7.12, Research, Estimates) all rose in tandem.

competes with

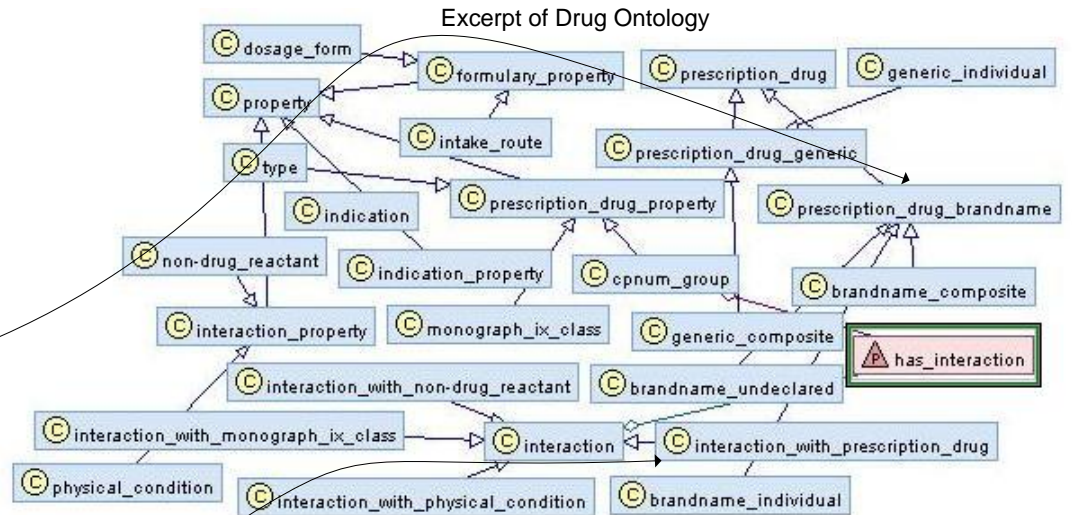
Semantic Annotation – news feed

<Entity id="494805" class="DrugOntology#Organization' Public Health Advisory FDA Announces Important Change
<Entity id="492805" class="DrugOntology#prescription_ Selective and Non-Selective Non-Steroidal Anti-Inflamma
class="DrugOntology#prescription_drug_generic">NSAI

Today, the Food and Drug Administration (FDA) is announ
<Entity id="474305" class="DrugOntology#Organization' to voluntarily withdraw <Entity id="122805" class=" DrugOntology#prescription_drug_brandname">Bextra
</Entity> (<Entity id="10288" class="DrugOntology#pre. valdecoxib</Entity>) from the market. Pfizer has agreed to of Bextra in the <Entity id="7852" class="Sweto#country" further discussions with the agency.

.....
This request is based on:
-Reports of serious and potentially life-threatening skin reactions, including deaths, in patients using Bextra. The risk of these reactions in individual patients is unpredictable, occurring in patients with and without a prior history of <Entity id="14280" class=" DrugOntology#interaction_with_physical_condition">sulfa allergy</Entity>

.....
Date created: <Regex type="date">April 7, 2005</Regex>

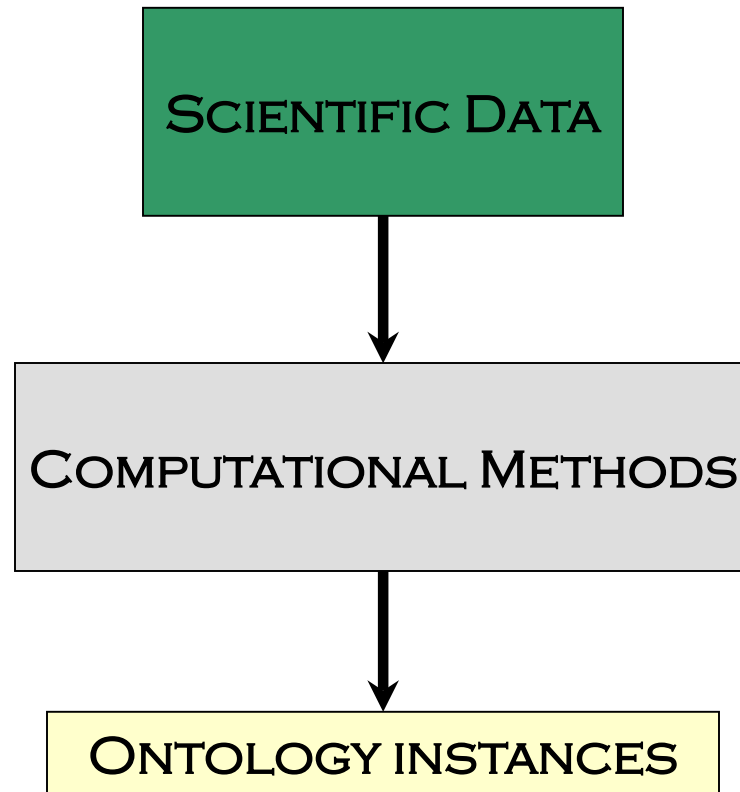


Sample Created Metadata

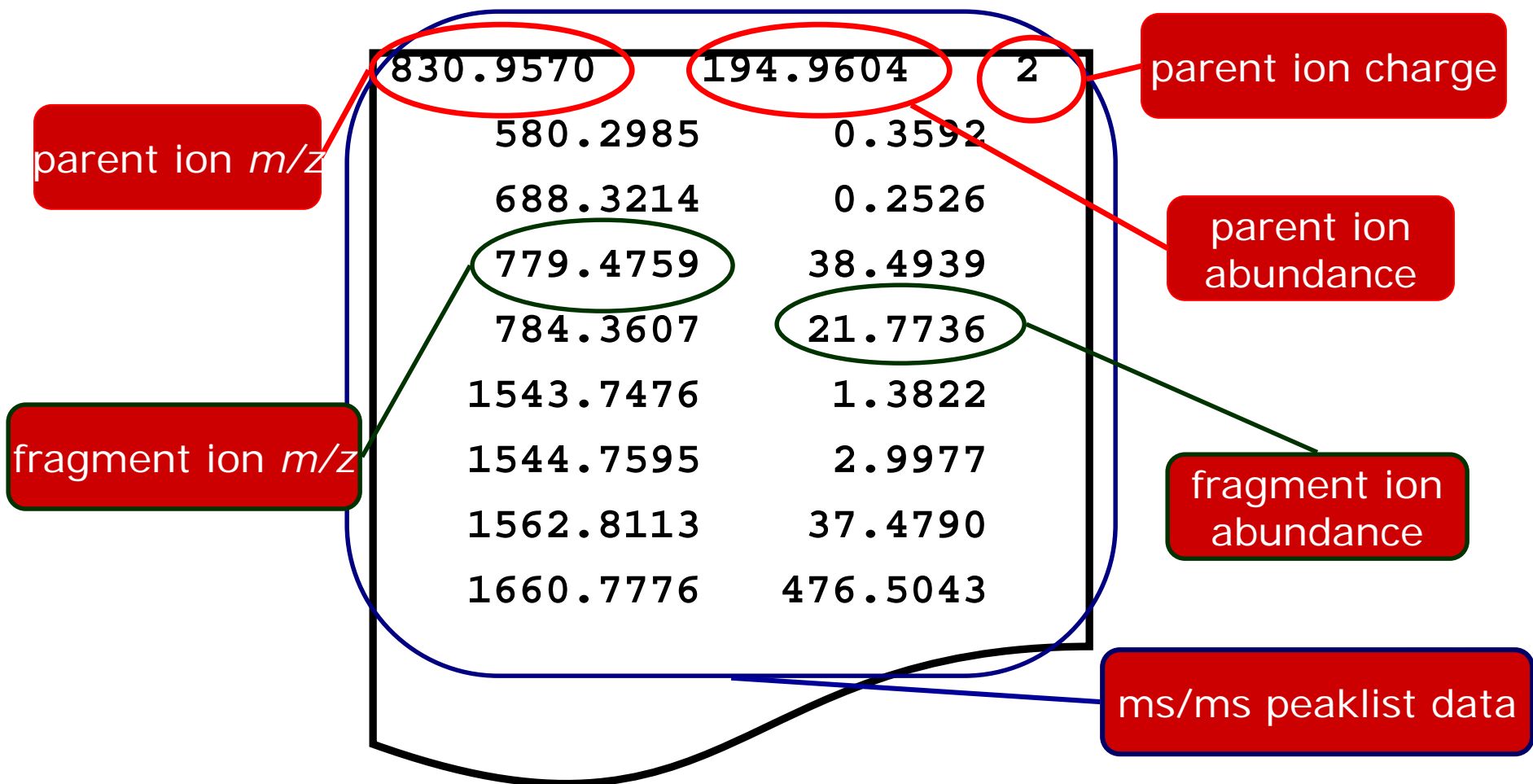
```
<Entity id="122805"
class="DrugOntology#prescription_drug_brandname">
Bextra
<Relationship id="442134"
class="DrugOntology#has_interaction">
<Entity id="14280" class="DrugOntology
#interaction_with_physical_condition">sulfa allergy
</Entity>
</Relationship>
</Entity>
```

Semantic annotation of scientific/experimental data

ProPreO population: transformation to rdf



ProPreO: Ontology-mediated provenance



MASS SPECTROMETRY (MS) DATA

<http://knoesis.org/research/bioinformatics>

PROPREO: ONTOLOGY-MEDIATED PROVENANCE

```
<ms-ms_peak_list>  
  <parameter instrument="micromass_QTOF_2_quadropole_time_of_flight_mass_spectrometer"  
  mode="ms-ms"/>  
  <parent_ion m-z="830.9570" abundance="194.9604" z="2"/>  
    <fragment_ion m-z="580.2985" abundance="0.3592"/>  
    <fragment_ion m-z="688.3214" abundance="0.2526"/>  
    <fragment_ion m-z="779.4759" abundance="38.4939"/>  
    <fragment_ion m-z="784.3607" abundance="21.7736"/>  
    <fragment_ion m-z="1543.7476" abundance="1.3822"/>  
    <fragment_ion m-z="1544.7595" abundance="2.9977"/>  
    <fragment_ion m-z="1562.8113" abundance="37.4790"/>  
    <fragment_ion m-z="1770.7777" abundance="476.5043"/>
```

Ontological
Concepts



SEMANTICALLY ANNOTATED MS DATA

Real World Applications

Application Types – by broad capabilities

- Search (also browsing, personalization)
- Integration (also interoperability)
- Analysis (also visualization)

Characterizing applications

Dimension: **Maturity**: Prototypes and demonstrations—show case research, technique

Systems using real world data, real <- our focus

Operational Systems <- our focus

Dimension: **Type of use**: General User, Consumer, Targeted Users/community on the Web, Business/Science/Engineering/Government User

Dimension: scalability -- Enterprise to Web

Dimension: realism Toy/synthetic data to **Real World Data**

Review of Applications

Ontology Development, Ontology Population Sources,
Ontology Quality Issues

Semantic Annotation/Metadata Extraction

Application Development Platforms/Support

Standards Usage

Scalability

Performance

Taalee's Semantic Search

Highly customizable, precise and freshest A/V search

Choose a category to search in: Music

Enter search term: Song

One field is usually enough.
Entering more yields more precise,
but fewer, results.

Artist: Madonna

Album/CD

Other Search Terms

Find me: Video Audio

Find It

Context and Domain Specific Attributes

Search results in Music for madonna
18 results found

Exact match for your search... **PRECISE SEARCH**

- 1. Material Girl (Madonna)**
ENGLISH SONGS FOR YOU ABCD'S
Source : Nawed Khan
Category: Music
Posted: 6/19/2000
Artist : Madonna
- 2. Holiday**
Madonna Louise Veronica Ciccone, 16 August 1958, Rochester Michigan, USA
Madonna excelled at dance and drama at high school...
Source : World Entertainment Network
Category: Music
Posted: 5/18/2000
Artist : Madonna
- 3. "Swim"**
Heard in background when Dawson is looking for Eve at the strip bar and the school office
Source : <http://www.dawsonscreenmusic.com/>
Category: Music
Posted: 4/19/2000
Artist : Madonna
- 4. Madonna - Evita**
Source : MTV
Category: Music
Posted: 4/14/2000
Artist : Madonna
- 5. The Power Of Good Bye**
Music video by madonna
Source : MTV
Category: Music
Posted: 4/14/2000
Artist : madonna
- 6. Nothing Really Matters**
Music video by madonna
Source : MTV
Category: Music
Posted: 1/09/2000
Artist : madonna

Uniform Metadata for Content from Multiple Sources, Can be sorted by any field

Rich Media Reference

Music RESULTS SEARCH

Holiday

Click to play
REAL

<http://shop.wen.com>

Madonna Louise Veronica Ciccone, 16 August 1958, Rochester Michigan, USA. Madonna excelled at dance and drama at high school and during brief periods at colleges in Michigan and North Carolina.

Produced by: World Entertainment Network
Posted Date: 5/18/2000
Artist: Madonna
Album: Immaculate Collection
Song Name: Holiday
Genre: Movies

Madonna Specials!	ORDER	Discography
American Pie DVD	\$32.99	Compilations
Immaculate Conception CD	\$13.99	Videography
Madonna Video Collection	\$17.99	Bibliography

[Email this to a friend](#)


Delightful, relevant information,
exceptional targeting opportunity


<http://www.streamingmediaworld.com/gen/reviews/searchassociation/>

What can a context do?

Microsoft Windows 2000 Professional only \$99.95 after mail-in rebates **CLICK HERE** SOFTWARE BUYLINE

(1) Business BACK TO LIST NEW SEARCH


Microsoft vows it will fight antitrust ruling 

(3)  <http://www.cbcnews.cbc.ca>

Microsoft Chairman Bill Gates is vowing to fight Monday's federal court decision that his company violated U.S. antitrust laws by abusing its monopoly power in personal computer operating systems.

Produced by: CBC News
Posted Date: 2000-04-04 12:00:00
Company Name: Microsoft Corp.
Ticker Symbol: MSFT
Exchange: NASDAQ

(2) Microsoft research by [HOOVERS](#)
Microsoft Corporation ([Nasdaq: MSFT](#)) [Capsule](#) [Financials](#) [Profile](#) [Officers](#)

 [Home](#) [shop](#) [help](#) [contact](#) [search](#)

Export Development Corporation
Accounts Receivable Insurance • Export Financing
www.edcinfo.com

RealPlayer: Microsoft vows it will fight antitrust ruling


File View Options Presets Sites Help

Channels

ADD NEW CHANNELS Subscribe To Free Channels

CNN.com
COMEDY CENTRAL
ABCNEWS.com

Updated 4:44 pm Update

Search 

Creating a Web of related information

Ontology driven Semantic Directory

The image shows a screenshot of a web browser displaying a Semantic Directory. The main page is titled "SEMANTIC DIRECTORY" and includes a search bar and navigation links. A red circle highlights the "Links" section, which lists various news sources and company profiles. A callout box points to the "Links" section, stating: "Links to news on companies that compete against". Another callout box points to the "Crucial news on Commerce One's competitors (Ariba) can be accessed easily and automatically".

SEMANTIC DIRECTORY
se-man-tic (sī-mān 'tīk) relating to meaning

Search where: Company Entire Directory

Relationships for Commerce One, Inc.

the following companyExecutives associated with the Commerce One, Inc.
[Hoffman, Mark P.](#)

the following companies competes with
[Ariba, Inc.](#)
[Catalyt Inc.](#)
[Click Commerce, Inc.](#)
[Metiom, Inc.](#)
[Yantra Corp.](#)
[eB2B Commerce, Inc.](#)
[Idapta, Inc.](#)
[Elcom International, Inc.](#)

the Commerce One, Inc. member(s) of
Corporate, Professional &

the Commerce One, Inc. identified by
[CMRC](#)

the Commerce One, Inc. competes with
[Ariba, Inc.](#)
[Oracle Corp.](#)

Other Frequently mentioned Company

- [Ariba, Inc.](#)
- [Prudential plc](#)
- [Centrica plc](#)
- [The IT Group, Inc.](#)
- [General Motors Corp.](#)
- [The Union Corp.](#)
- [First Union Corp.](#)
- [VerticalNet, Inc.](#)
- [SAP AG](#)
- [The Dow Corning](#)

Links hide **Company**

Sites of Interest:
[On24](#)
[FT News](#)
[Bloombergs Television](#)
[Bloomberg](#)
[Medialink Business](#)
[StartDustTalk Radio](#)
[CNET Radio](#)
[CNNFN The Financial Network](#)
[FT.com archive](#)
[DallasBusiness](#)
[Hovers Company Data](#)
[CNNfn The Financial Network](#)
[CNNfn The Financial Network Digital Jam](#)
[dbusiness](#)
[OhioBusinessEducation](#)
[CBS MarketWatch](#)
[Executive Rad](#)
[CATO](#)
[CNNFNBriefs](#)
[CNBC Video Highlights](#)

Ariba, Inc. via Company

See [Relationships](#)

Audio/Video clips for Ariba, Inc.
Pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [Next](#) ([Next 10 pages](#)) of 24

- Ariba, Inc.**
ON The Move: B2B Marketplace Stocks Slump on Downgrade
Source : ON24 Posted: 3/8/2001
Category: Business Company Name : Ariba, Inc.
- Analyst Cuts Estimates for B2B and E-business Software...**
Thomas Weisel Partners Analyst David Gremmels downgrades Ariba, i2 and Siebel from Strong Buy to Buy and keeps his Market Perform...
- Analyst Cuts Estimates for B2B and E-business Software...**
Thomas Weisel Partners Analyst David Gremmels downgrades Ariba, i2 and Siebel from Strong Buy to Buy and keeps his Market Perform...

Semantic Targeting

What else can a context do?
(a commercial perspective)

Baseball RESULTS SEARCH

Clemens injured

ONE-CLICK MEDIA PLAY

Click to play audio

REAL

<http://www.espn.com>

Roger Clemens did the best he could to get out of the first inning, taken out after strained right groin.

Produced by: ESPN
Posted Date: 6/14/2000
League: MLB
Teams: New York Yankees
Players: Roger Clemens

City and Team not mentioned in story. Taalee Knowledge Experts added these. Other Searchers for 'Yankees' would not find this story.

Baseball stats from MajorLeagueBaseball.com

[Clemens Stats](#)
[Yankees Stats](#)
[Yankees Schedule](#)
[Buy Yankee Tickets](#)

Category specific sponsorship dynamically added by Taalee.

Semantic Enrichment

Semantic/Interactive Targeting

The screenshot illustrates a web interface for semantic targeting. On the left, a search bar is labeled "Enter search term" and "One t". Below it, "Search Tips" and "Choose another c" are visible. The main content area is titled "Movies" and features "The Insider" by Russell Crowe. It includes a "Click to play video" section with buttons for "WINDOWS", "REAL", "QUICKTIME", and "REAL". A yellow arrow points from the search bar to the movie page, and a blue arrow points from the movie page to the CDNOW product listing. The CDNOW listing shows "The Insider" (1999) with a price of \$26.39 (DVD) and \$15.99 (VHS), and "Donnie Brasco" (1997) with a price of \$20.97 (DVD) and \$9.49 (VHS). A blue banner at the bottom reads: "Precisely targeted through the use of Structured Metadata and integration from multiple sources".

Enter search term

One t

Entering more yields more prec

Search Tips

Choose another c

Movies

The Insider

Click to play video

WINDOWS

REAL

QUICKTIME

WINDOWS

REAL

http://www.hollywood.com

Russell Crowe stars as Jeffrey Wigand, the real-life executive on "60 Minutes" unveils the tobacco industry's cover-up of the smoking. Al Pacino and Christopher Plummer play Lowell Bee Wallace, the venerable producer and journalist who are press story by the network. Based on a Vanity Fair article by Marie

Produced by: Hollywood

Posted Date: 3/29/2000

Genre: Drama

Cast: Al Pacino, Russell Crowe, Christopher Plummer, Diane Venora

Directed by Michael Mann. Written by Marie Brenner, Eric Rot

Credits: Based on the article by Marie Brenner. Produced by Pieter Jan Mann. Released by Buena Vista.

Rating: R, for language

Film Title: The Insider

Buy Al Pacino Videos

Buy Russell Crowe Videos

Buy Christopher Plummer Videos

Buy Diane Venora Videos

Buy Philip Baker Hall Videos

Buy The Insider Video

CDNOW

Never miss a beat.™

Shopping Cart contains 0 items

Al Pacino

Shop

- Music
- Videos

Articles & Media

- Reviews

Who & What

- Biography

Insider

1999: Starring Actor

*Release Date: Nov. 7

Up to 20% Off ↓

DVD \$26.39

*VHS \$15.99

Any Given Sunday

1999: Starring Actor

20% Off ↓

DVD \$19.98

Al Pacino Collection

1999

VHS \$44.99

Donnie Brasco

1997: Starring Actor

Up to 30% Off ↓

*DVD \$20.97

VHS \$9.49

Internet

Semantic Application – Equity Dashboard



Automatic 3rd party content integration

Focused relevant content organized by topic (semantic categorization)

Related news not specifically asked for (Semantic Associations)

Automatic Content Aggregation from multiple content providers and feeds

Competitive research inferred automatically

Semantic Applications: Health Care

Active Semantic Medical Records

(operational since January 2006)

Goals:

- Increase efficiency
- Reduce Errors, Improve Patient Satisfaction & Reporting
- Improve Profitability (better billing)

Technologies:

- Ontologies, semantic annotations & rules
- Service Oriented Architecture

Thanks -- Dr. Agrawal, Dr. Wingeth, and others. [Active Semantic Electronic Medical Record ISWC2006 paper](#)
http://iswc2006.semanticweb.org/items/in_use_8.php

Semantic Web Applications in Government

Passenger Threat Analysis

Need to Know -> Demo

Financial Irregularity *

* a classified application

Primary Funding by ARDA, Secondary Funding by NSF
An Ontological Approach to the Document Access Problem of Insider Threat

Semantic Application in a Global Bank

Aim: Legislation (PATRIOT ACT) requires banks to identify 'who' they are doing business with

Problem

Volume of internal and external data needed to be accessed

Complex name matching and disambiguation criteria

Requirement to 'risk score' certain attributes of this data

Approach

Creation of a 'risk ontology' populated from trusted sources (OFAC etc); Sophisticated entity disambiguation

Semantic querying, Rules specification & processing

Solution

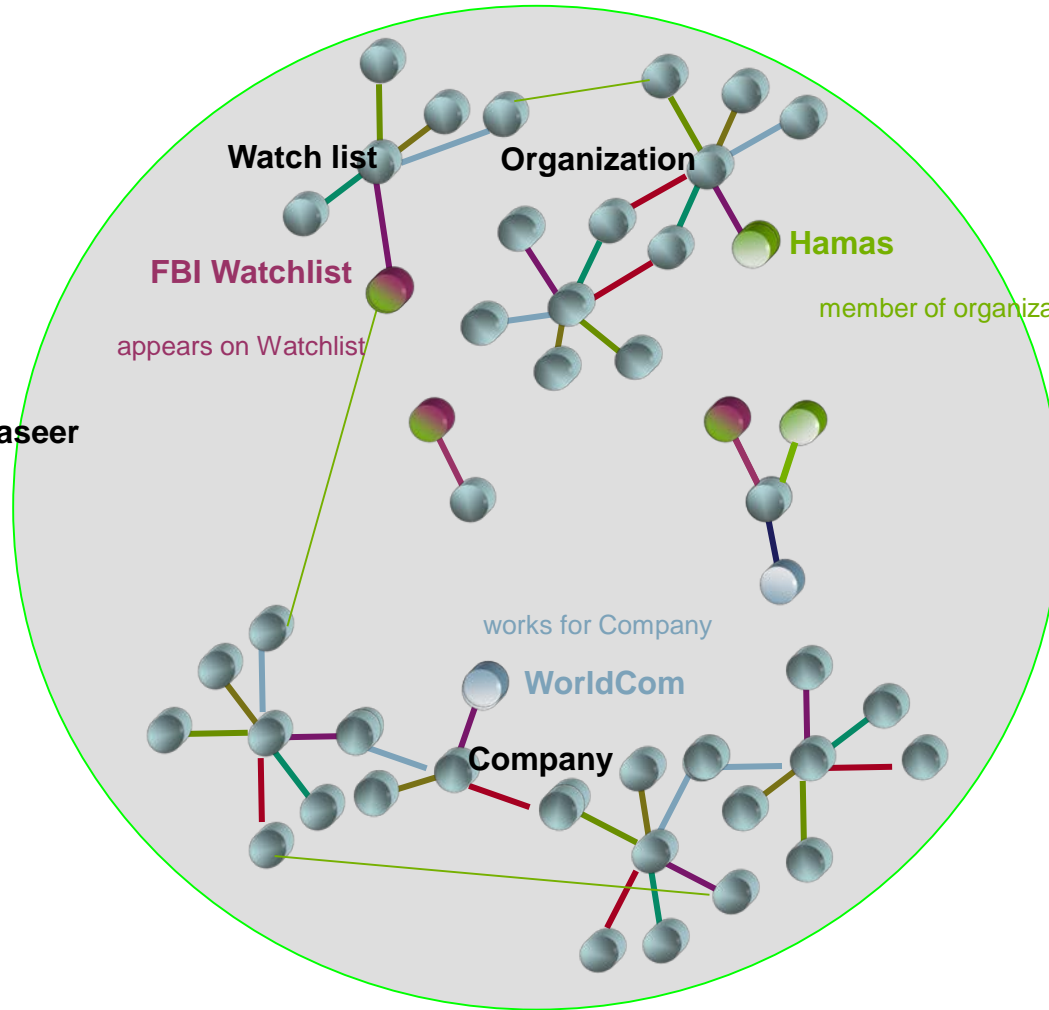
Rapid and accurate KYC checks

Risk scoring of relationships allowing for prioritisation of results; Full visibility of sources and trustworthiness

The Process

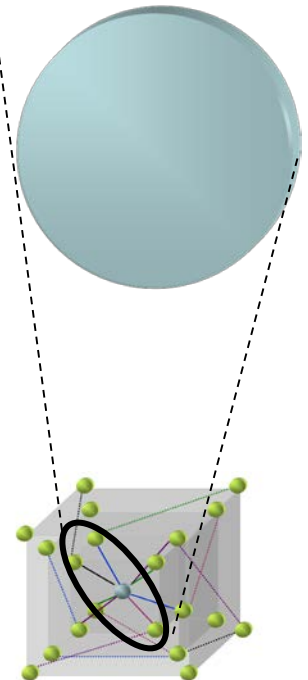


Ahmed Yaseer

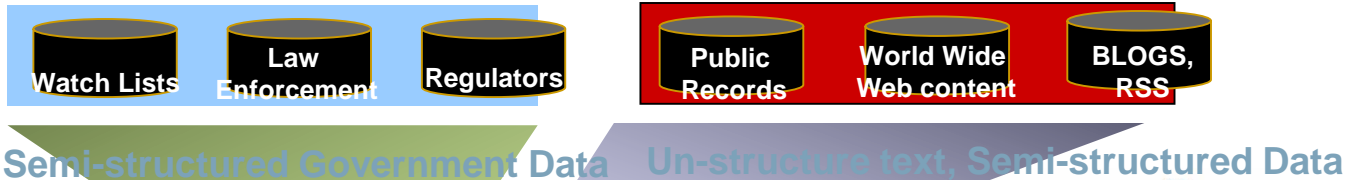


Ahmed Yaseer:

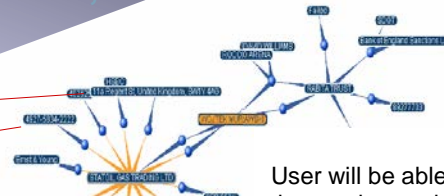
- *Appears on Watchlist 'FBI'*
- *Works for Company 'WorldCom'*
- *Member of organization 'Hamas'*



Global Investment Bank



Establishing New Account



User will be able to navigate

Example of Fraud Prevention application used in financial services

Individual Check - Microsoft Internet Explorer

File Edit View Favorites Tools Help

CIRAS
Customer Identification and Risk Assessment

User: Larry Parker | Sign Off

Check Individual | **Check Organization** | Status & Results | Reporting

LMZ Shipping, LLC ENTITY Score: 65 | Print (PDF)

Check ID: 14092 Status: Pending — Cancel or Refer this Record

Summary | Match Details | WebFountain | Media Content | Attributes | **Associations**

Launch Associations Visualizer

Is Related to	Wojtech Moroski Wojtech Moroski is related to Rabbita Trust Rabbita Trust appears on FBI Watch List
Undertakes	12 Thompson Rd, Albany, Georgia
Is related to	Wojtech Moroski
Active in	Washington
Active in	US

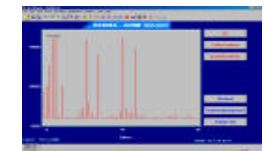
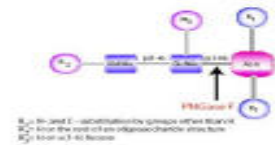
Perform a New Check

SEMAGIX
POWER • THROUGH • RELEVANCE

[Terms, conditions, caveats and small print](#)

© Copyright 2004. Semagic. All Rights Reserved.

N-Glycosylation Process (NGP)

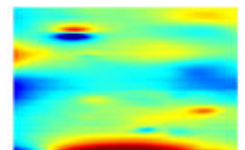
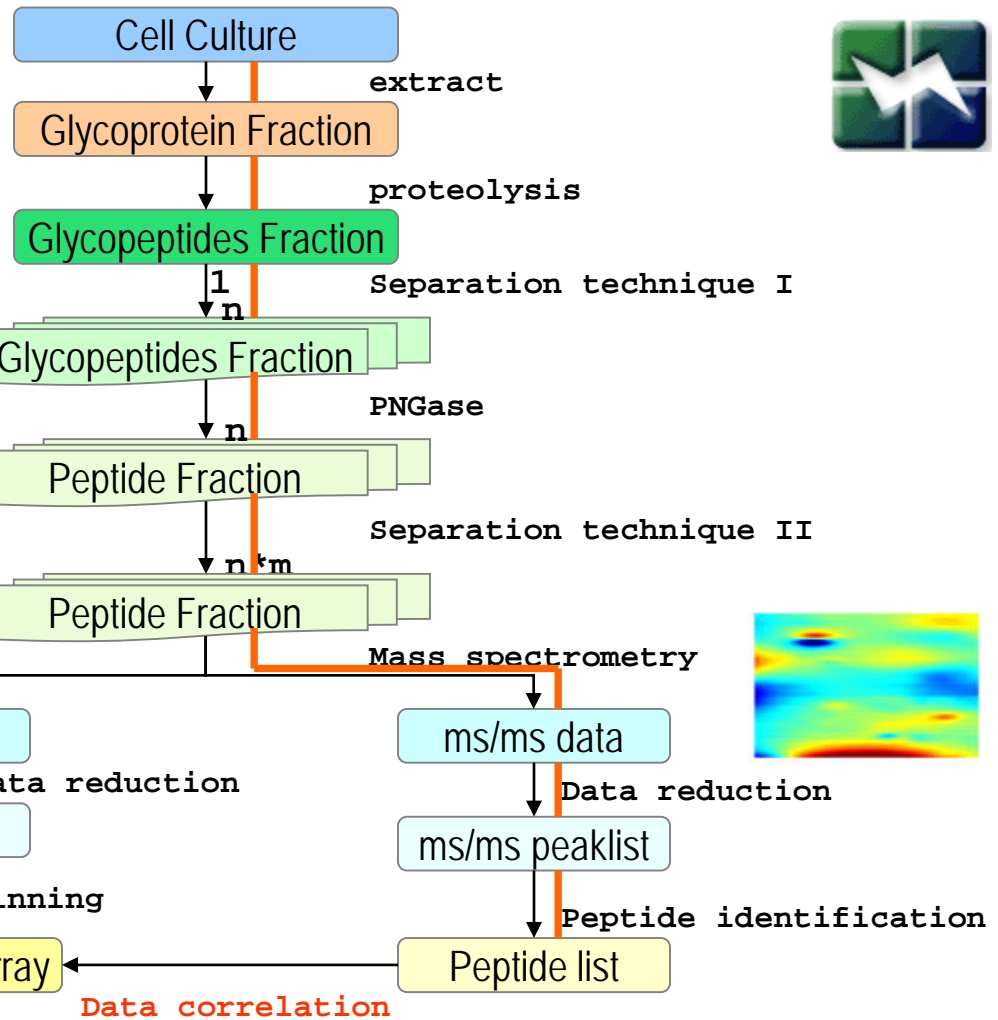


Glycopeptide identification and quantification

Signal integration

N-dimensional array

Data correlation



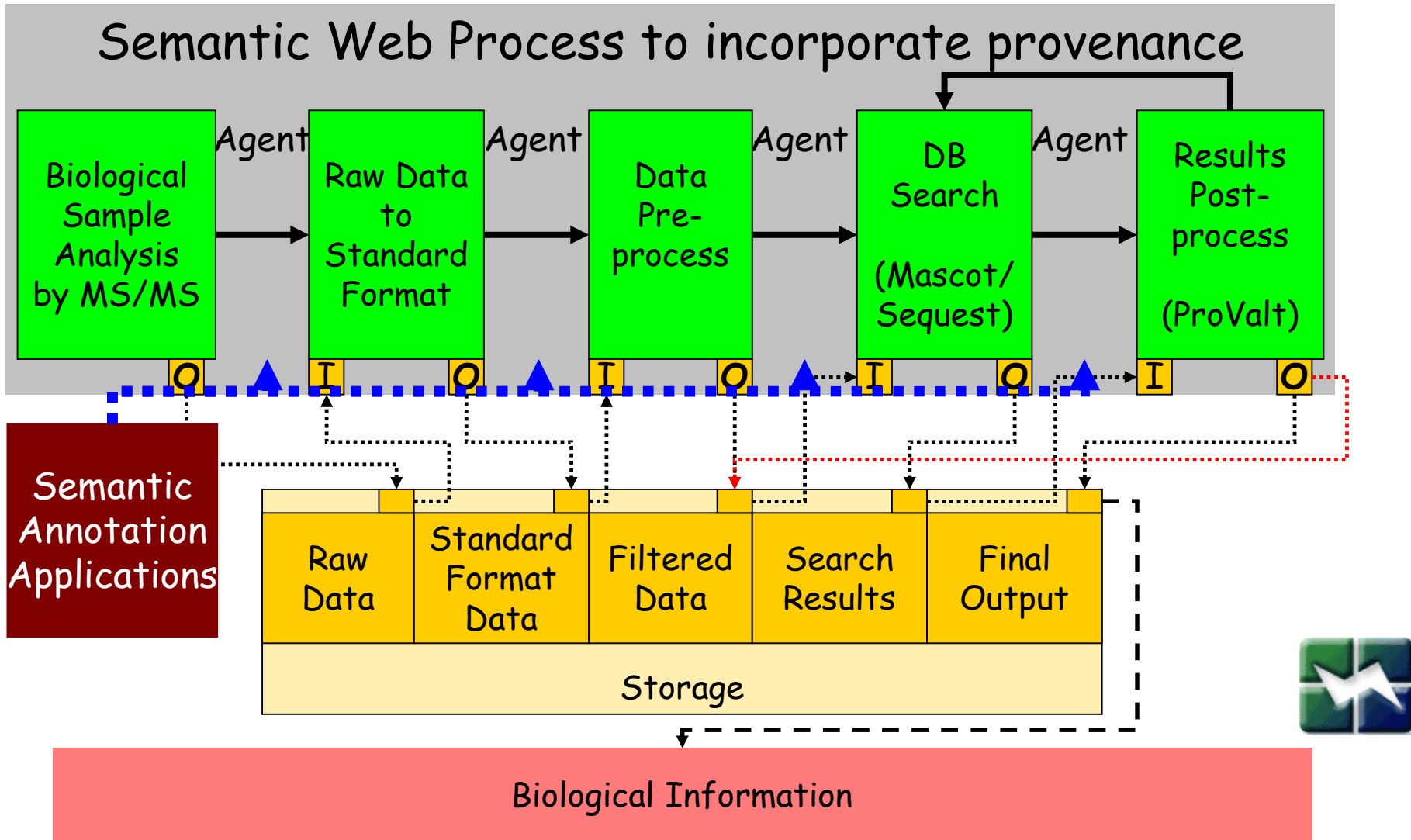
Workflow based on Web Services = Web Process

The screenshot displays the Scuff Workbench v1.5-APSHOT interface, which is used for configuring and running workflows based on web services. The interface is divided into several panels:

- Tools and Workflow Invocation:** Located at the top left, it features the myGrid logo and the text "Taverna Scuff Workbench v1.4".
- Run Workflow Panel:** This panel allows users to manage workflow inputs. It includes a "Load Inputs" section with a tree view showing a hierarchy of folders and files (e.g., "user", "satya", "folder", "F:\James\RawDemo\raw", "p", "do_mascot?", "true"). Below the tree is a "Load" section with options for "Load" and "Load from URL". A "Run Workflow" button is located at the bottom right of this panel. A text box below the panel provides instructions: "To input data into this workflow you must select the item from the tree to the left of this panel and either enter the data manually, upload from a file on your local machine or load from a location on the internet. When all workflow inputs have been populated as required you can click the Run workflow button to run the workflow on these inputs."
- Workflow Diagram Panel:** This panel shows a visual representation of the workflow. It includes a "Save as" section and a "Configure diagram" button. The diagram itself is a flowchart with various nodes representing tasks and decision points. The nodes are connected by arrows indicating the flow of the process. The diagram is rendered and ready for execution.
- Processor Status Panel:** This panel provides a detailed view of the workflow's execution. It includes a "Status" tab and a "Process report" section. The "Process report" section contains a table with columns for "Type", "Name", "Last event", "Event timestamp", "Event detail", and "Breakpoint". Below the table is a "Graph" section showing the workflow's progress, with nodes for "Inputs", "Intermediate inputs", "Intermediate outputs", and "Outputs". The "Inputs" section shows the "do_mascot?" node, which branches into "conditionalT" and "conditionalF". The "Intermediate inputs" section shows nodes like "initFolderPath", "massSpecReq", "rawTomzXML", "storeFile_samba", and "sendFile_samba". The "Intermediate outputs" section shows nodes like "mzXMLtoPKI", "pSplit", "normal_Mascot", "storFile_normal_Mascot", "folderPathMod", "random_Mascot", "sendFile_random_Mascot", and "storeFile_random_Mascot". The "Outputs" section shows the "proValtSearch" node.

The bottom of the screenshot shows the system tray with the date and time: "Tue Mar 6, 2:20 PM".

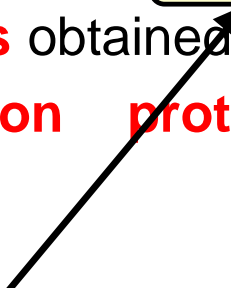
ISiS – Integrated Semantic Information and Knowledge System



Semantic Annotation Facilitates Complex Queries

- *Evaluate the specific effects of changing a biological parameter:* Retrieve **abundance** data for a given **protein** expressed by three different **cell types** of a specific **organism**.
- *Retrieve raw data supporting a structural assignment:* Find all the **raw ms data files** that contain the **spectrum** of a given **peptide sequence** having a specific **modification** and **charge state**.
- *Detect errors:* Find and **compare** all **peptide** lists identified in **Mascot output files** obtained using a similar **organism, cell-type, sample preparation protocol**, and **mass spectrometry** conditions.

A Web Service
Must Be Invoked



ProPreO concepts highlighted in red

More Life Science Applications

- Semantic Browser: contextual browsing of PubMed

More at [Knoesis research in life sciences](#)

Relationship Extraction

About the data used

UMLS – A high level schema of the biomedical domain

- 136 classes and 49 relationships
- Synonyms of all relationship – using variant lookup (tools from NLM)

T147—effect
T147—induce
T147—etiology
T147—cause
T147—effecting
T147—induced

MeSH

- Terms already asserted as instance of one or more classes in UMLS

PubMed

- Abstracts annotated with one or more MeSH terms

Example PubMed abstract (for the domain expert)

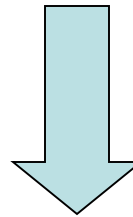
```
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Display&DB=pubmed
IP - 13
DP - 2006 Apr 13
TI - [Catheter ablation of atrial fibrillation]
PG - 40-3; quiz 44
AB - In patients with drug-refractory atrial fibrillation, left-atrial catheter
ablation represents a new curative therapeutic option. Segmental ostial or
circumferential pulmonary vein isolation can achieve stable sinus rhythm
in some 70% of patients with paroxysmal atrial fibrillation but no severe
structural heart disease. In patients with chronic atrial fibrillation,
complex left-atrial linear, or substrate-oriented ablation strategies may
additionally be applied. In patients with cardiac insufficiency or more
severe systolic left-ventricular dysfunction, restoration of a stable
sinus rhythm through the use of left-atrial catheter ablation can improve
the left-ventricular ejection fraction and reduce the severity of cardiac
failure. Potential complications of ablation include, in particular,
pulmonary veins stenosis, iatrogenic left-atrial tachycardia,
thromboembolic events and fatal atrio-esophageal fistulas.
AD - Medizinische Klinik I, Klinikum Grosshadern, Universitat Munchen.
Christopher.Reithmann@med.uni-muenchen.de
FAU - Reithmann, C
AU - Reithmann C
FAU - Remp, T
AU - Remp T
LA - ger
PT - Journal Article
PT - Review
TI - Katheterablation bei Vorhofflimmern. Läsionen errichten elektrische
Blockaden.
PL - Germany
TA - MMW Fortschr Med
JT - MMW Fortschritte der Medizin.
JID - 100893959
SB - IM
MH - Atrial Fibrillation/physiopathology/*surgery
MH - Atrial Flutter/etiology
MH - *Catheter Ablation/adverse effects
MH - Chronic Disease
MH - English Abstract
MH - Esophageal Fistula/etiology
MH - Heart Failure, Congestive
MH - Humans
MH - Iatrogenic Disease
MH - Stroke Volume
MH - Tachycardia/etiology
MH - Thromboembolism/etiology
MH - Ventricular Dysfunction, Left
RF - 0
EDAT- 2006/05/23 09:00
MHDA- 2006/06/13 09:00
PST - ppublsh
SO - MMW Fortschr Med. 2006 Apr 13;148(15):40-3; quiz 44.
```

Abstract

Classification/Annotation

Method – Parse Sentences in PubMed

[1254239-1] An excessive endogenous or exogenous stimulation by estrogen induces adenomatous hyperplasia of the endometrium.

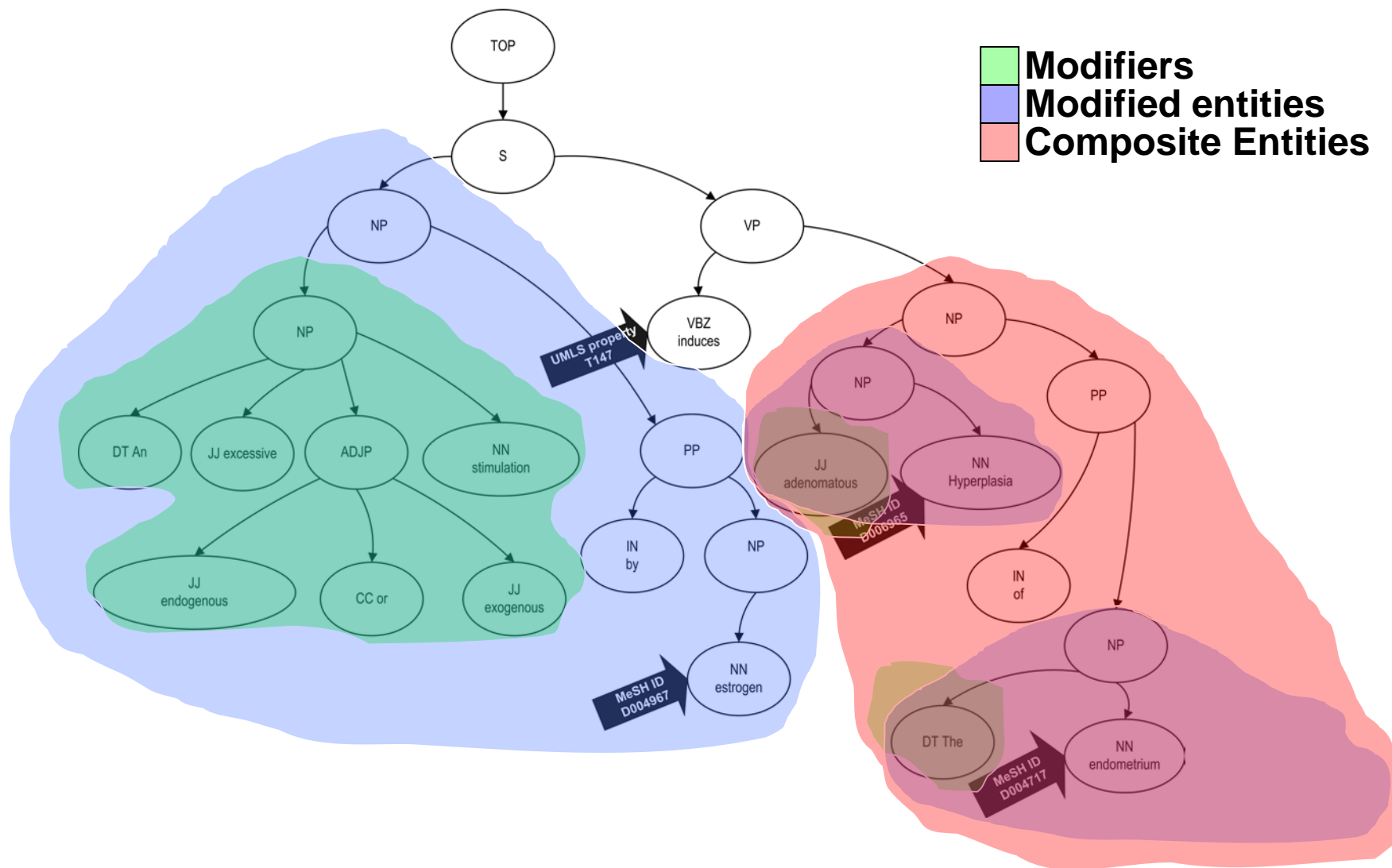


SS-Tagger (University of Tokyo)

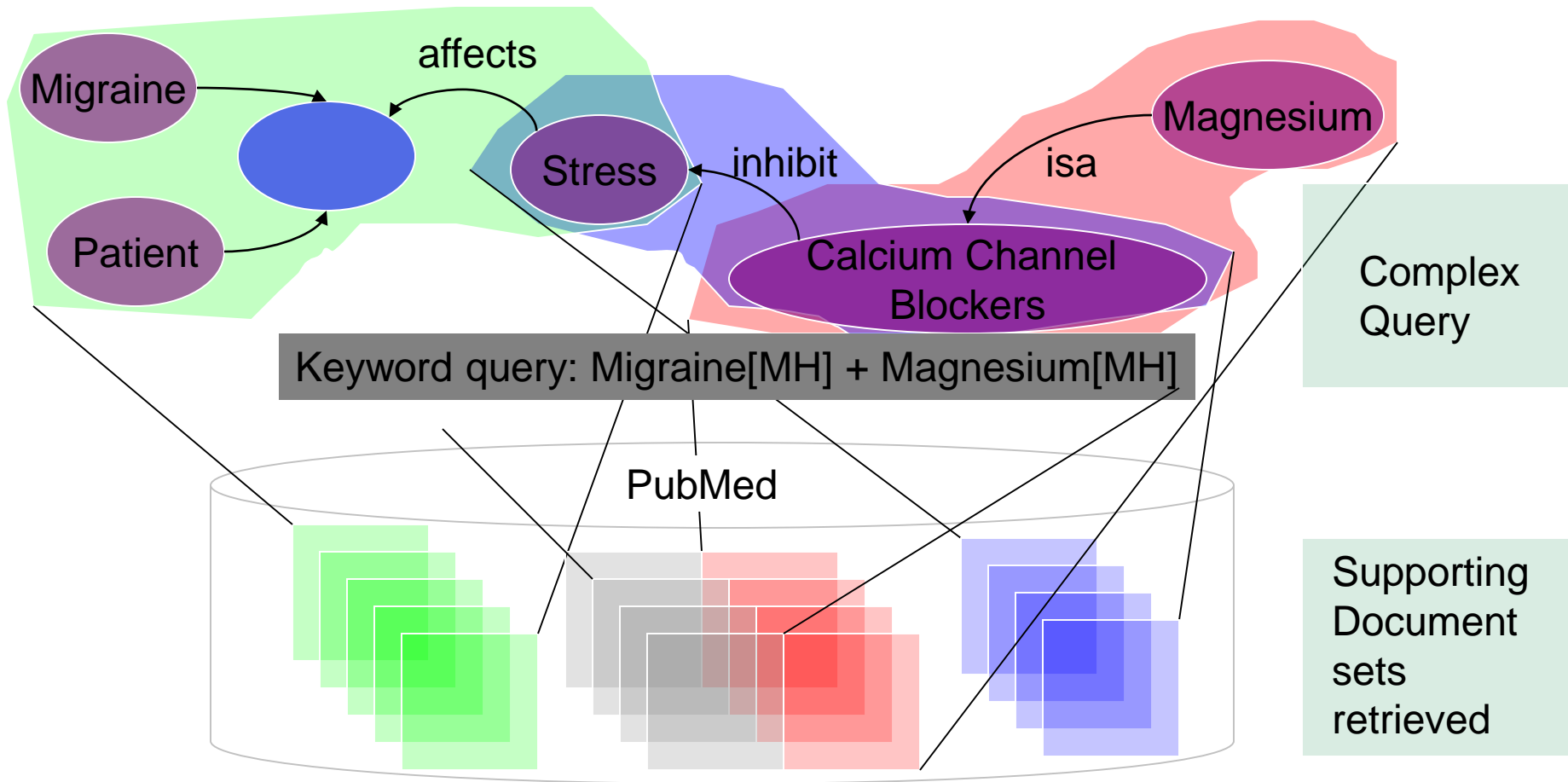
SS-Parser (University of Tokyo)

(TOP (S (NP (NP (DT An) (JJ excessive) (ADJP (JJ endogenous) (CC or) (JJ exogenous)) (NN stimulation)) (PP (IN by) (NP (NN estrogen)))) (VP (VBZ induces) (NP (NP (JJ adenomatous) (NN hyperplasia)) (PP (IN of) (NP (DT the) (NN endometrium)))))))))))

Method – Identify entities and Relationships in Parse Tree



Hypothesis Driven retrieval of Scientific Literature



Eli Lilly Case Study

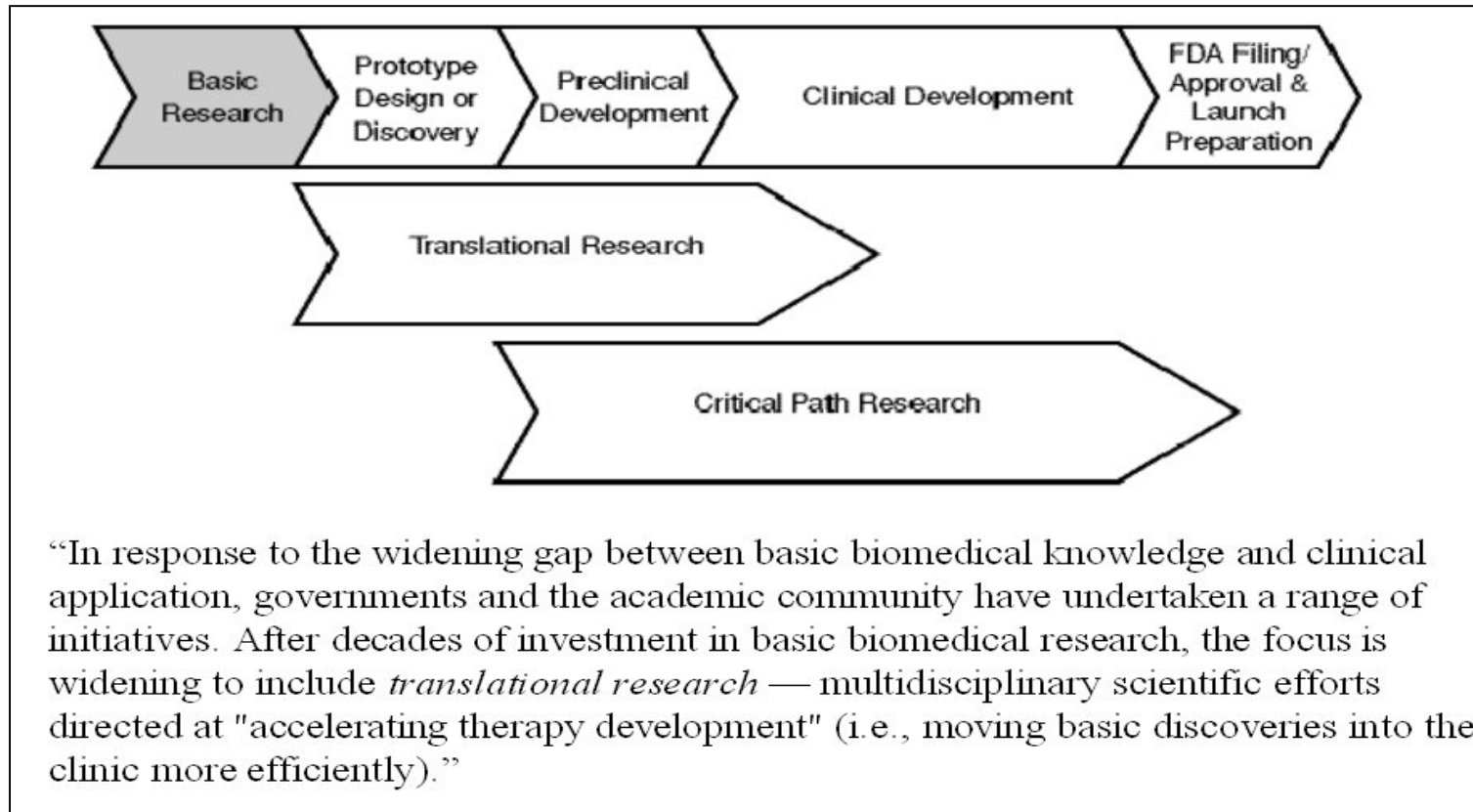
Advances in Science

- Tremendous advances in biology over the last decade
 - Sequencing of the human genome
 - Technology for large scale expression studies and patient genotyping
 - High resolution imaging (e.g. whole organism, cell based)
 - Discovery and development of siRNA and other techniques
- Facilitates a more tailored approach to therapeutics

Tailored Therapeutics

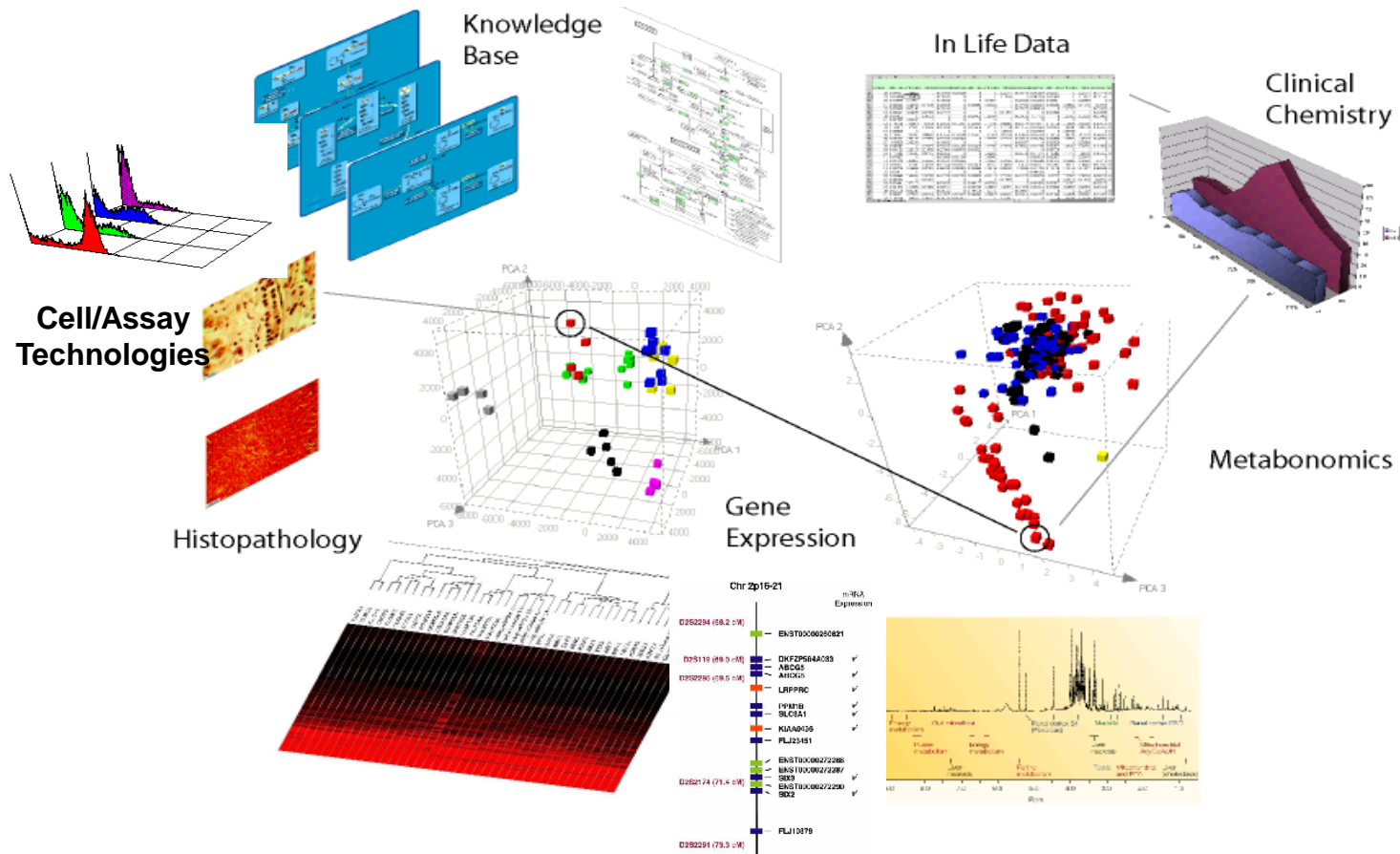
- The end of the blockbuster era
- The right dose to the right patient at the right time
- Lowering cost of pharmaceutical development
- Improved patient response to medication

Critical Path Initiative

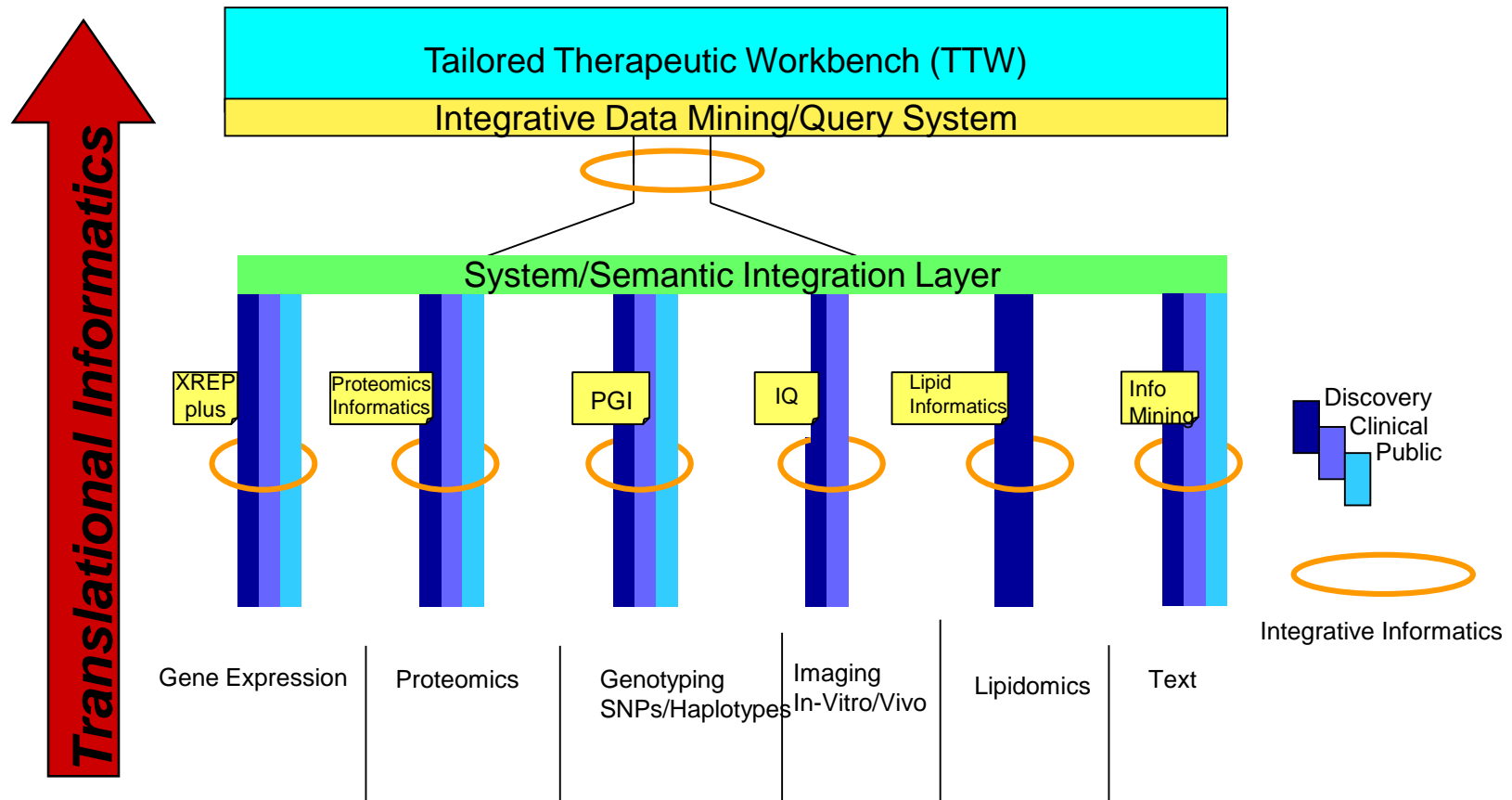


Source: Innovation or Stagnation, FDA Report, March 2004

The Web of Heterogeneous Data



Integrative Informatics Platform



Discovery Target Assessment Tool

The screenshot displays the Lilly Science Grid Discovery Target Assessment Tool interface, which is used for analyzing and assessing potential drug targets. The interface is divided into several main sections:

- Gene Information (Top Left):** Displays details for the gene **SCN5A** (Entrez Gene ID: 6331). It includes the species (Homo sapiens), synonyms (e.g., hML1, hML2, hML1-NV, LQT3, SBB1, CDCD2, CVD1E, CVD2, Nav1.5), description (sodium channel, voltage-gated, type V, alpha subunit), summary (protein encoded by this gene is an integral membrane protein and tetrodotoxin-resistant voltage-gated sodium channel subunit), phenotypic type (MIM: 601144; Brugada syndrome), chromosome (3 Map: 3q21), and organism (Homo sapiens; human; man).
- Ontology (Middle Left):** Shows a hierarchical tree of ontologies for **SCN5A**, including Competitors, Competitors Stage, MESH, Diseases (Cardiovascular Diseases, Heart Diseases, Arrhythmia), and MeSH.
- Available TINS data for CDK1/CDC2 (Bottom Left):** A table listing various data points for the target, such as Organism (HUMAN), Gene Chromosome Map (10), Gene Type (protein-coding), and Gene Status (REVIEWED). It also lists associated diseases like Ovarian Neoplasms, Multiple Myeloma, and various Leukemias.
- Gene Detail (Bottom Right):** Provides a comprehensive overview of the **Scn5a** gene, including its symbol, name (voltage channel, voltage-gated, type V, alpha), synonyms (hML1, Nav1.5, Nav1.5C, SMI1), gene structure (70.0 kb), and sequence map (Chr1:19332109-19427714 bp). It also lists representative sequences and their lengths.
- Target Intelligence (Top Right):** A network diagram showing relationships between the target and other entities, including various proteins and pathways.
- Genomic Map (Middle Right):** A visual representation of the gene's location on Chromosome 3, showing exons and introns.

Use Cases for RDF

- Patient Stratification
 - If we could 'semantically' describe a patient with biomarker/genomic properties, we could 'semantically' compare them. New definition of *similarity*.
- Cellomics Data
 - If we could 'semantically' describe cellular localization with other properties, we could discover novel indicators (e.g. cell size relates to expression) - *emergent properties*.
- Exon Descriptors
 - If we could 'semantically' describe all the exons on a genome and relate transcription and SNPs to other functional consequences, we could 'reason' across the genome (i.e. query in interesting ways). Adding power to *interrogation*.

Summary

- The life sciences industry needs to increase productivity
- The Semantic Web is a promising technology for data integration and decision support
- Lilly is using Semantic Web technologies for the Target Assessment Tool within drug discovery

Acknowledgements

John Reynders, Ph.D, Discovery Informatics

Alan Palkowitz, Ph.D., Discovery Chemistry

Jim Stephens, Ph.D., Discovery Toxicology

Jude Onyia, Ph.D., Discovery Biology

Harry Harlow, Ph.D., Integrative Biology

Dan Robertson, Ph.D., Computational Chemistry

Discovery Target Assessment (D-TAT) Team

Semantic Web in Practice

Agenda

- Tools
- Data / vocabularies
- Collateral
- Community
- Pointers for getting going

Semantic Web Infrastructure

- Triple stores
- RDFizers
- Ontology Editors / Reasoning Systems
- Application Frameworks

Triple Stores

- 3Store
- Aduna
- AllegroGraph
- Boca
- Joseki
- Kowari
- Mulgara
- Oracle RDF Data Model
- Profium Metadata Server
- RDF Gateway
- RDFStore
- Sesame
- Virtuoso
- YARS

There are many others available too...

Boca, IBM

IBM Internet Technology IBM

Boca Overview – Multi-user, distributed enterprise RDF repository

- Selective RDF replication from server to client machines
- Security, including named-graph-based RDF access control
- Audit trails of changes to data within named graphs
- Near real-time event notifications
- Sophisticated programming model

```
graph TD; DB2[(DB2)] --- RS[Repository Server]; RS -- Updates --> UM((Update Manager)); subgraph Application; direction LR; RM[Remote Model]; LM[Local Model]; end; Application --- C[(Cache)]; UM -- Updates --> Application; WS[Web Service] --> Application;
```

The diagram illustrates the Boca architecture. At the top, a 'Repository Server' is connected to a 'DB2' database. The 'Repository Server' sends 'Updates' to an 'Update Manager' (represented by a cloud). The 'Update Manager' then sends 'Updates' to the 'Application' layer. The 'Application' layer consists of a 'Remote Model' and a 'Local Model'. A 'Cache' is also connected to the 'Application' layer. A 'Web Service' is shown interacting with the 'Application' layer.

Features of an Enterprise-ready Semantic Web Storage System – SiCoP 2007 © 2007 IBM Corporation

Source: IBM

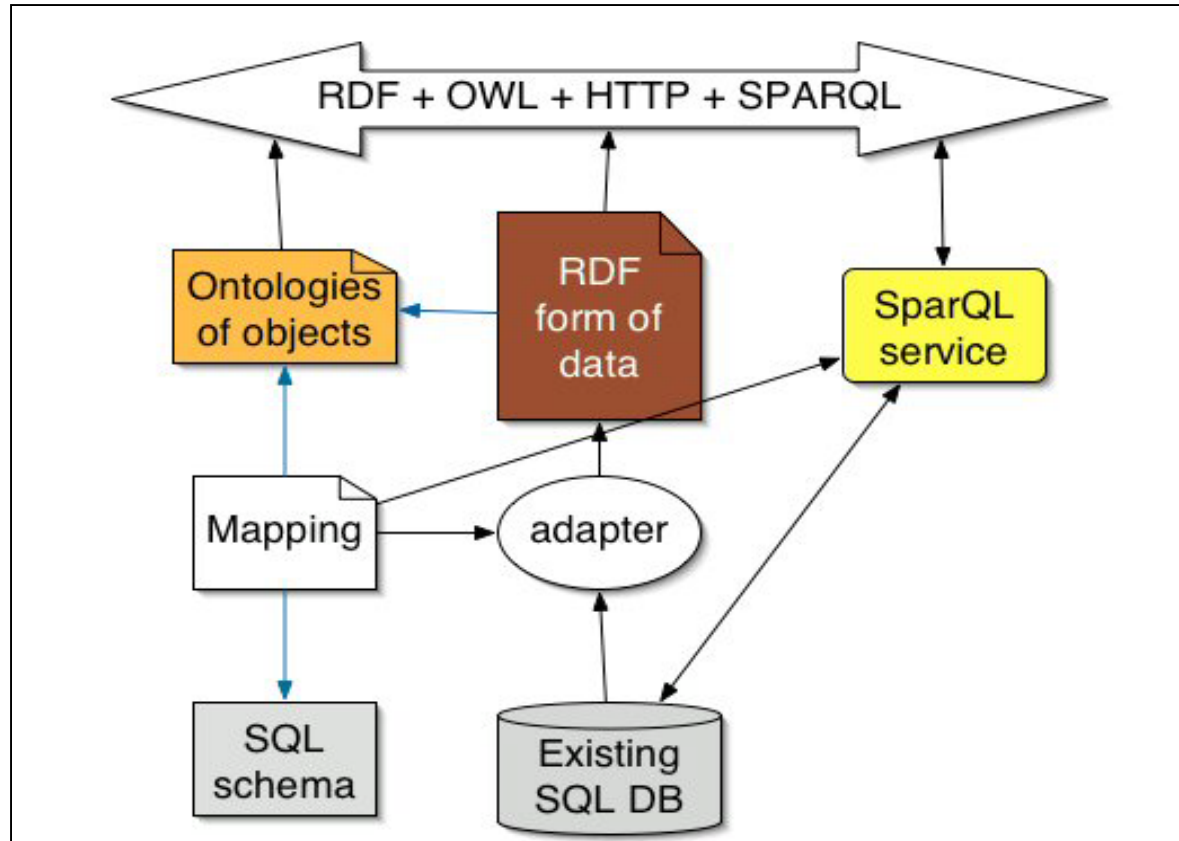
RDF Data Model, Oracle

- Object-relational implementation
- Set of triples form an RDF/OWL graph (model)
- Optimized storage structure: repeated values stored only once
- Can handle multiple lexical forms of the same value
- Incremental load and bulk load
- SPARQL-like graph pattern embedded in SQL query
- Native inferencing for RDF, RDFS & user-defined rules
- Support for OWL and Semantic Operators in the next release

Virtuoso, OpenLink

- Hybrid Data Server that combines SQL, RDF, XML, and Full Text Data Management
- Includes a Virtual / Federated DBMS Layer that enables transparent access to data from 3rd party SQL, RDF, XML, and Web Services
- Produces RDF Instance Data in Physical and Virtual forms from local or 3rd party data sources
- Provides full support for the SPARQL Query Language against Physical and Virtual RDF Graphs
- Query Optimizer is specifically tuned for high-performance data access across all realms
- Includes in-built middleware for producing RDF instance data on-the-fly from non RDF Data Sources (e.g. (X)HTML, Microformats, Web Services, Binary Files)

Adapting SQL Databases



Source: Tim Berners-Lee

Mapping Relational to RDF

- D2RQ
- SquirrelRDF
- DartGrid
- SPASQL

The screenshot displays the Dart User Toolkits interface with several panels and annotations:

- 1. Display relational tables:** A callout box pointing to the 'Tables Tree View' on the left, which lists various tables including 'ZT_VIEW_ZHEJIANLIF'.
- 2. Display ontologies:** A callout box pointing to the 'DataSchema Tree View' at the bottom left, which shows a hierarchical structure of ontologies.
- 3. User drags tables and classes, and establishes their mappings:** A callout box pointing to the central workspace where two ontologies, 'ZT_VIEW_ZHEJIANLIF' and 'ZT_VIEW_ANNMOLF', are displayed. Red arrows indicate the mapping between the tables in the 'Tables Tree View' and the properties in these ontologies.
- 4. Outline of the mapping definitions:** A callout box pointing to the 'Outline' panel on the right, which shows a list of mapping definitions for the selected table.
- 5. Meta-information about the selected table:** A callout box pointing to the 'Properties' panel at the bottom, which displays details for the selected table, including its URI and schema.

Source: DartGrid

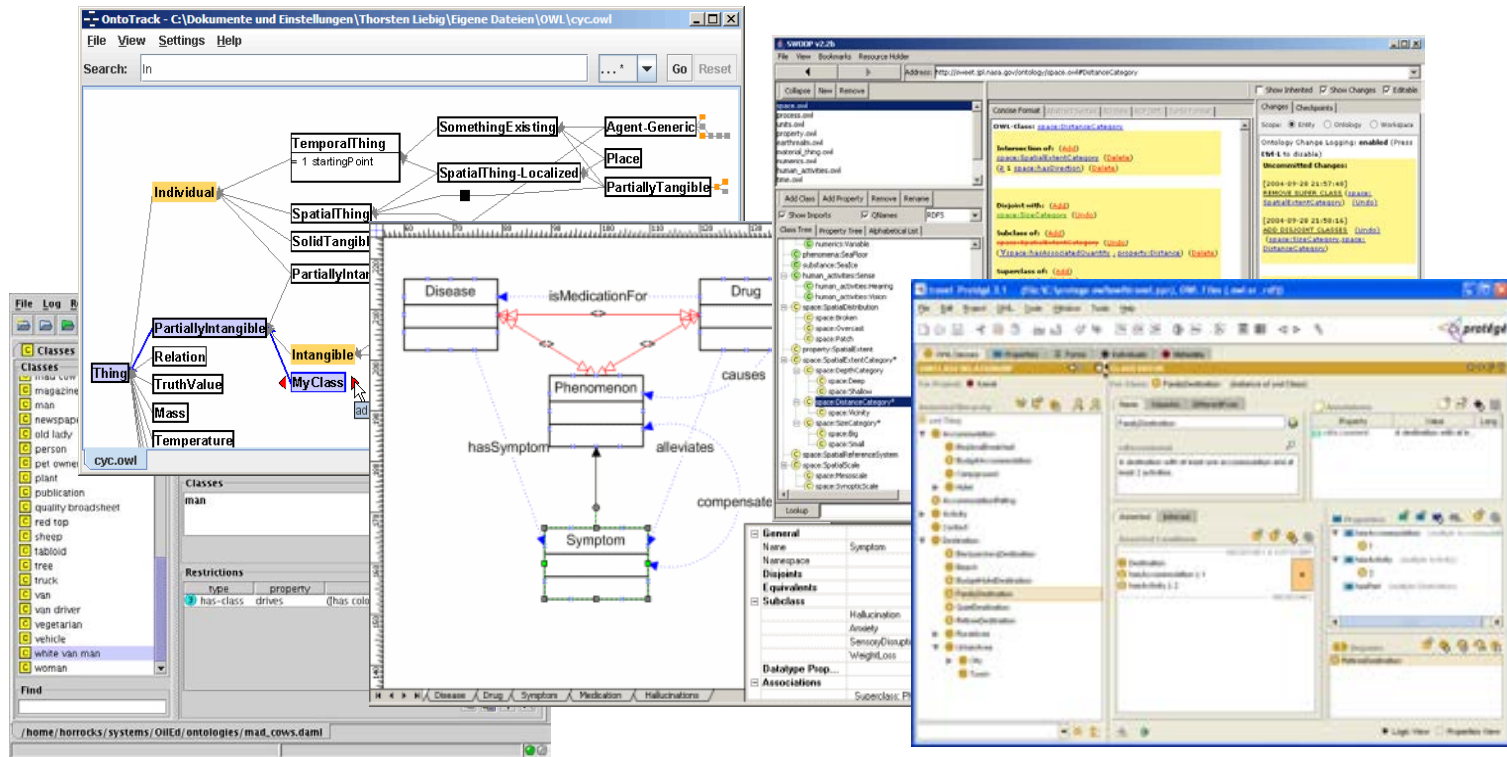
RDFizers

- Relational -> RDF
- XML -> RDF
- Excel -> RDF
- JPEG -> RDF
- BibTEX -> RDF
- Java -> RDF
- Weather -> RDF
- Palm -> RDF
- Outlook -> RDF
- Flickr -> RDF

A directory of RDFizers is provided at:
<http://simile.mit.edu/wiki/RDFizers>

Ontology Editors and Environments

Protégé, SWOOP, GrOWL, TopBraid, Ontotrack, SemanticWorks, ..



Source: Ian Horrocks

Reasoning Systems



FaCT++



Pellet



Semantic Web Tools

- RDF programming environment for 14+ languages
 - C, C++, C# and .Net, Haskell, Java, Javascript, Lisp, Obj-C, PHP, Perl, Prolog, Python, Ruby, Tcl/Tk
- Selection of on-line validators
 - BBN OWL Validator, OWL Consistency Checker, WonderWeb OWL-DL Validator, RDF Validator, RDF/XML & N3 Validator, ConsVISor OWL Consistency Checker
- SPARQL Endpoints
 - SPARQLer, SPARQLette, XML Army Knife, OpenLink Virtuoso
- Semantic Web Crawlers
 - Swoogle, SWSE, Zitgist

Semantic Web Tools

- RDF Browsers
 - BrowseRDF, /facet, Longwell, mSpace, Siderean Software, Exhibit
- Semantic Web Browsers
 - DISCO, ObjectViewer, OpenLink RDF Browser, Tabulator Browser, Haystack
- Labeling
 - Adobe XMP
- Information Extraction
 - Amilcare, Language and Computing
- Visualization
 - IsaViz, Perfuse, Tom Sawyer, RDF-Gravity
- Relationship Analytics
 - Cogito
- Content Management
 - Profium Semantic Information Router
- Information Integration
 - Ontoprise, Software AG, @Semantics, webMethods, Revelytix, Ontology Works

Over 500 tools are now available

Lists of Tools

- <http://sites.wiwiss.fu-berlin.de/suhl/bizer/toolkits/index.htm>
- <http://esw.w3.org/topic/SemanticWebTools>
- <http://www.mkbergman.com/?p=291>
- <http://planetrdf.com/guide/>
- http://www.sekt-project.org/resources/sekt_components.html

How to get RDF Data?

- Write your own RDF in your preferred syntax
- Add RDF to XML directly (in its own namespace), e.g. in SVG
- Use intelligent scrapers or wrappers to extract RDF from a Web page and then generate automatically (e.g. via an XSLT script)
- Formalize the scraper approach with GRDDL
- RDFa extend (X)HTML by defining general attributes to add metadata to any element
- Create bridge to relational databases
- Use bridge from other data sources

RDF Data

- Annotea Bookmark File
- DBLP
- dbpedia
- dbtune
- Geonames
- MusicBrainz
- RDF Book Mashup
- Revyu
- US Census Data
- WordNet
- BIND
- BrainPharm
- Entrez Gene
- HIVSDB
- KEGG
- NeuroNames
- Reactome
- SenseLab
- SWAN publication & hypothesis
- UniProt

Vocabularies

- eClassOwl: eBusiness ontology for products and services
- Gene Ontology: describes gene and gene products
- BioPAX: for biological pathway data
- SKOS core: describes knowledge systems, thesauri, glossaries
- Dublin Core: about information resources, digital libraries, with extensions for rights, permissions, digital rights management
- FOAF: about people and their organizations
- DOAP: on the descriptions of software products
- Music Ontology: describes CDs, music tracks, etc.
- SIOC: for semantically-Interlinked Online Communities

Source: Ivan Herman

Collateral

- Much good information at W3C
 - <http://www.w3.org/2001/sw/>
- New FAQ on the Semantic Web
 - <http://www.w3.org/2001/sw/SW-FAQ>
- Semantic Web Case Studies and Use Cases
 - <http://www.w3.org/2001/sw/sweo/public/UseCases>
- List of Semantic Web books
 - <http://esw.w3.org/topic/SwBooks>
- Dave Beckett's Resources
- PlanetRDF a blog aggregator on Semantic Web topics

Public Fora at W3C

- Semantic Web Interest Group
 - A forum for developers with an archived mailing list, and a constant IRC presence on [freenode.net#swig](https://freenode.net/#swig)
- Semantic Web for Health Care & Life Sciences: SW-HCLS
- Semantic Web Deployment Working Group
 - Archives of working group are public
- Semantic Web Education and Outreach IG
 - Community Projects
 - Whitelisting Email Senders with FOAF
 - Linking Open Data on the Semantic Web
 - Knowee Contact Organizer
 - POWDER Browser Extension

Pointers for Getting Going

- Use robust URIs
- Reuse existing data and ontologies
- A little semantics goes a long way
- Model the real world rather than data artifacts
- Build upon your infrastructure incrementally

Summary

- Many Semantic Web tools are available
- Data and vocabularies are increasingly being made available in RDF/OWL
- Many books, tutorial and overviews are available to help you get going
- Several public fora for community activities

Copy of the tutorial will be available at:

<http://knoesis.wright.edu> -> Library