Wright State University

CORE Scholar

Kno.e.sis Publications

The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis)

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

Semantic Web: Technologies and Applications for the Real-World

Susie Stephens
Principal Research Scientist

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

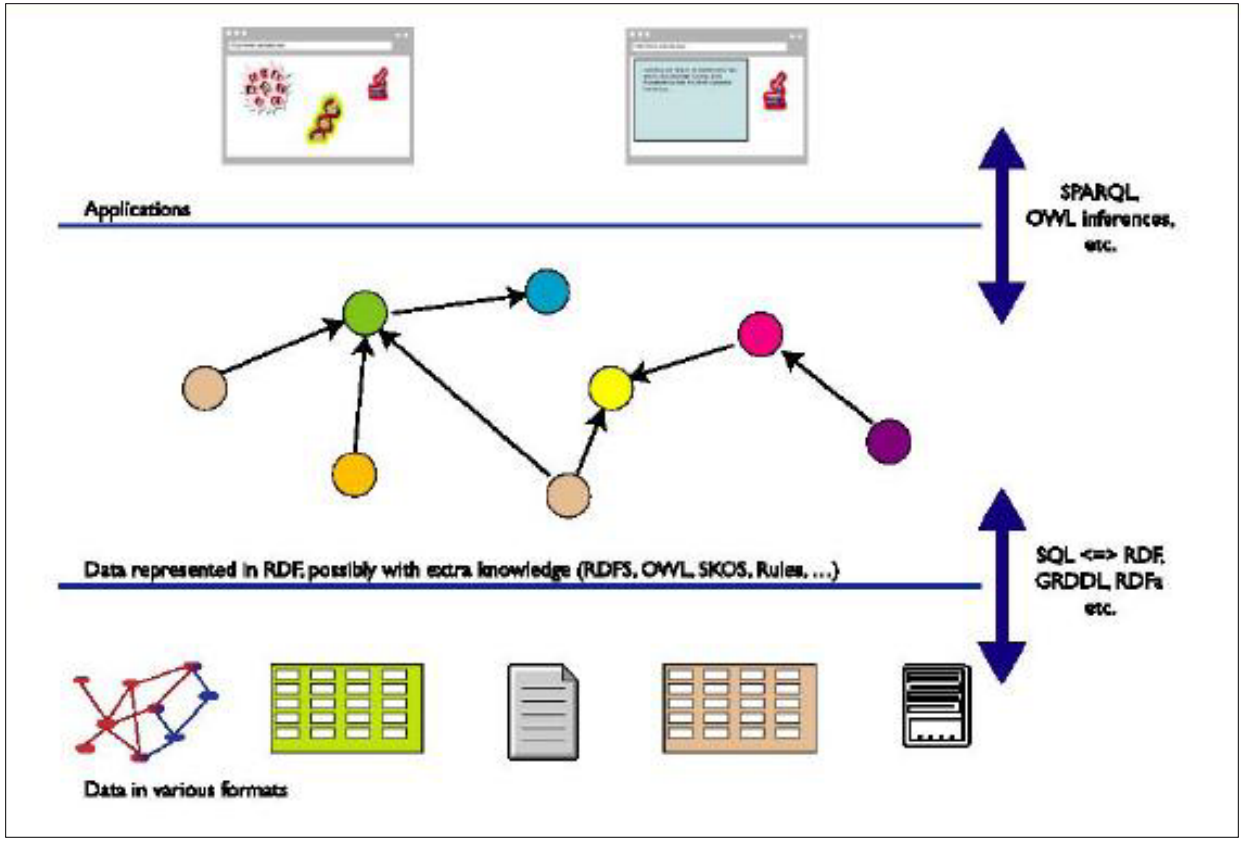
Agenda

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

Characterizing the Semantic Web

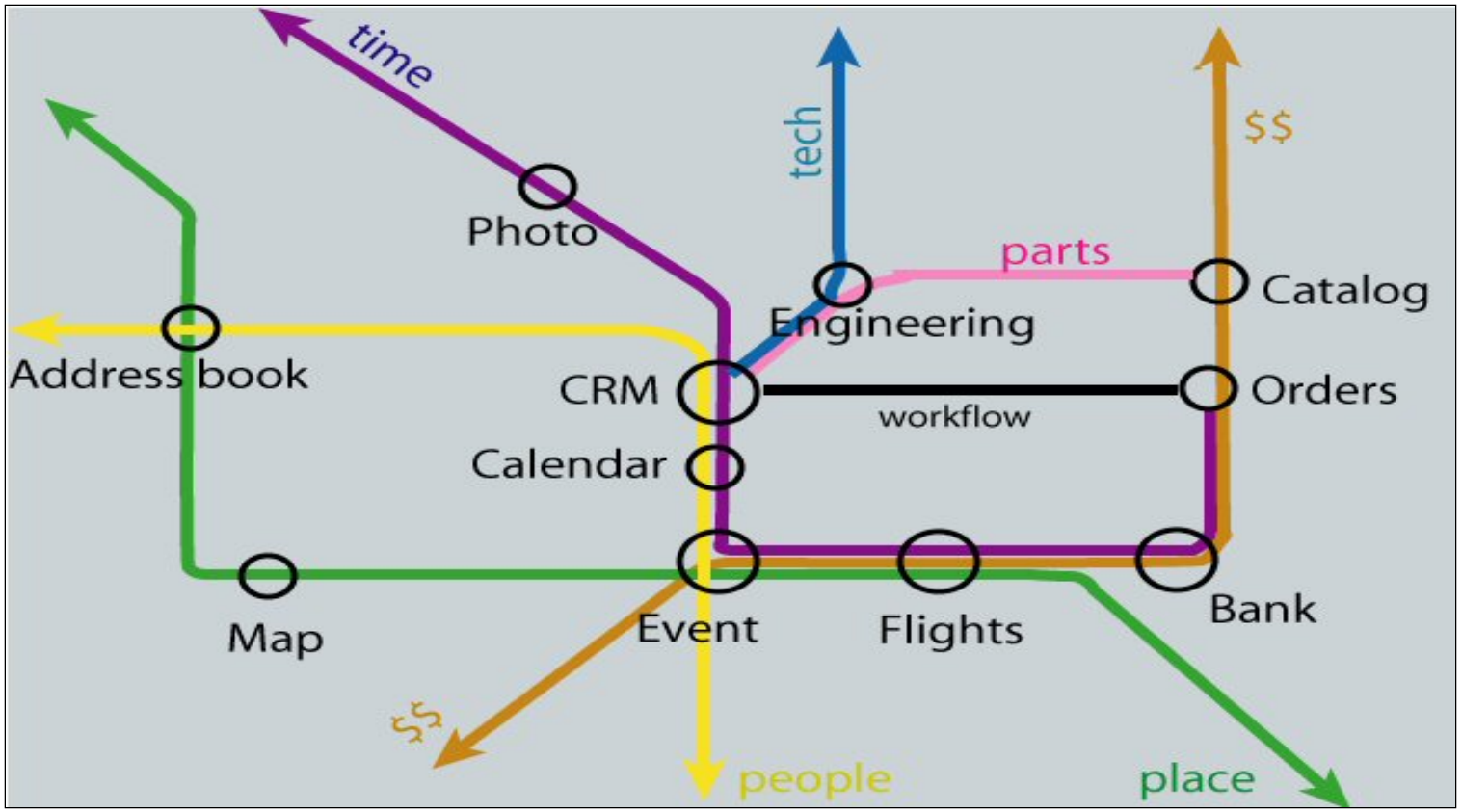
- Semantic Web is an interoperability technology
- An architecture for interconnected communities and vocabularies
- A set of interoperable standards for knowledge exchange

Creating a Web of Data



Source: Ivan Herman

Mashing Data

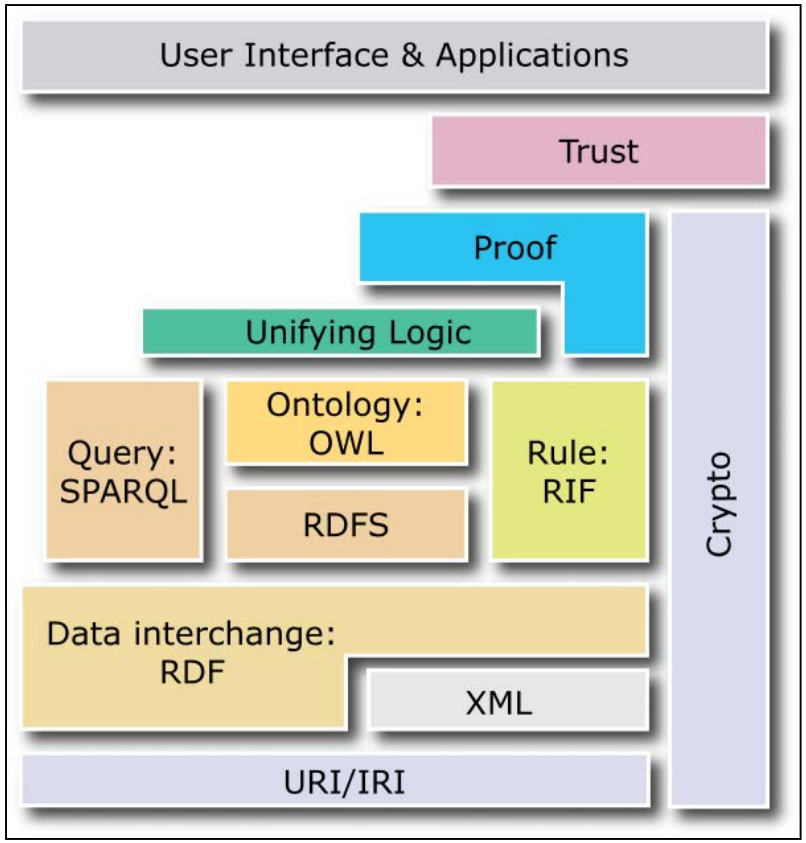


Source: W3C

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

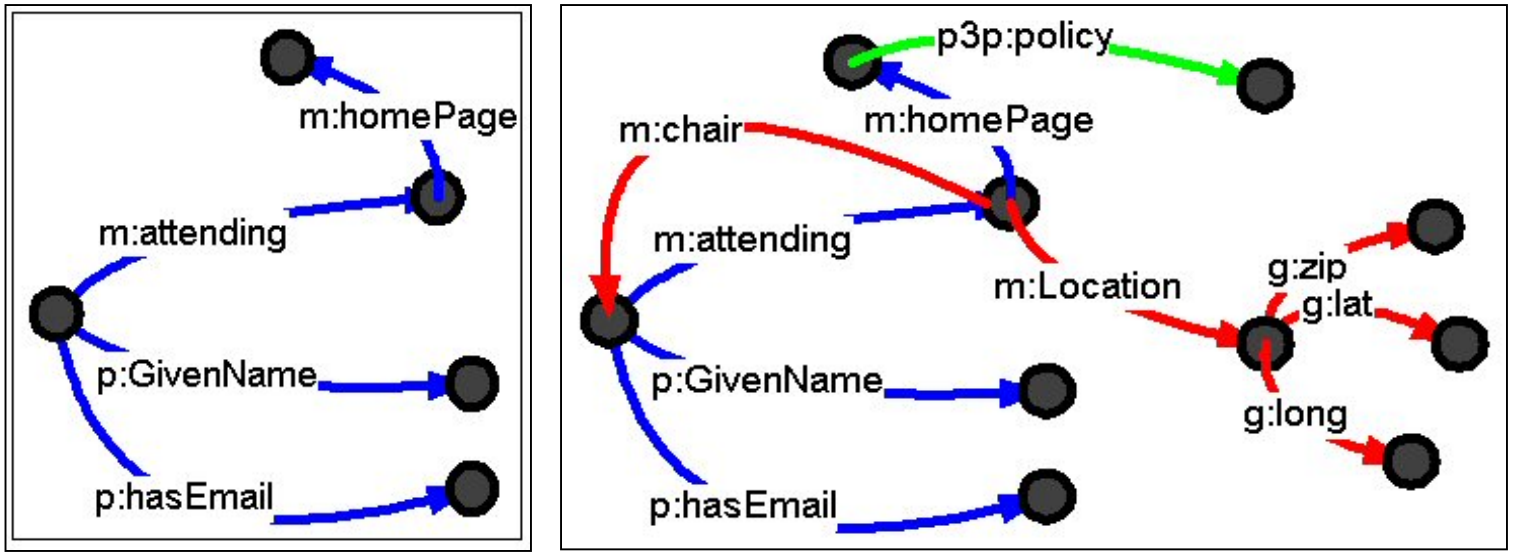


Source: W3C

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

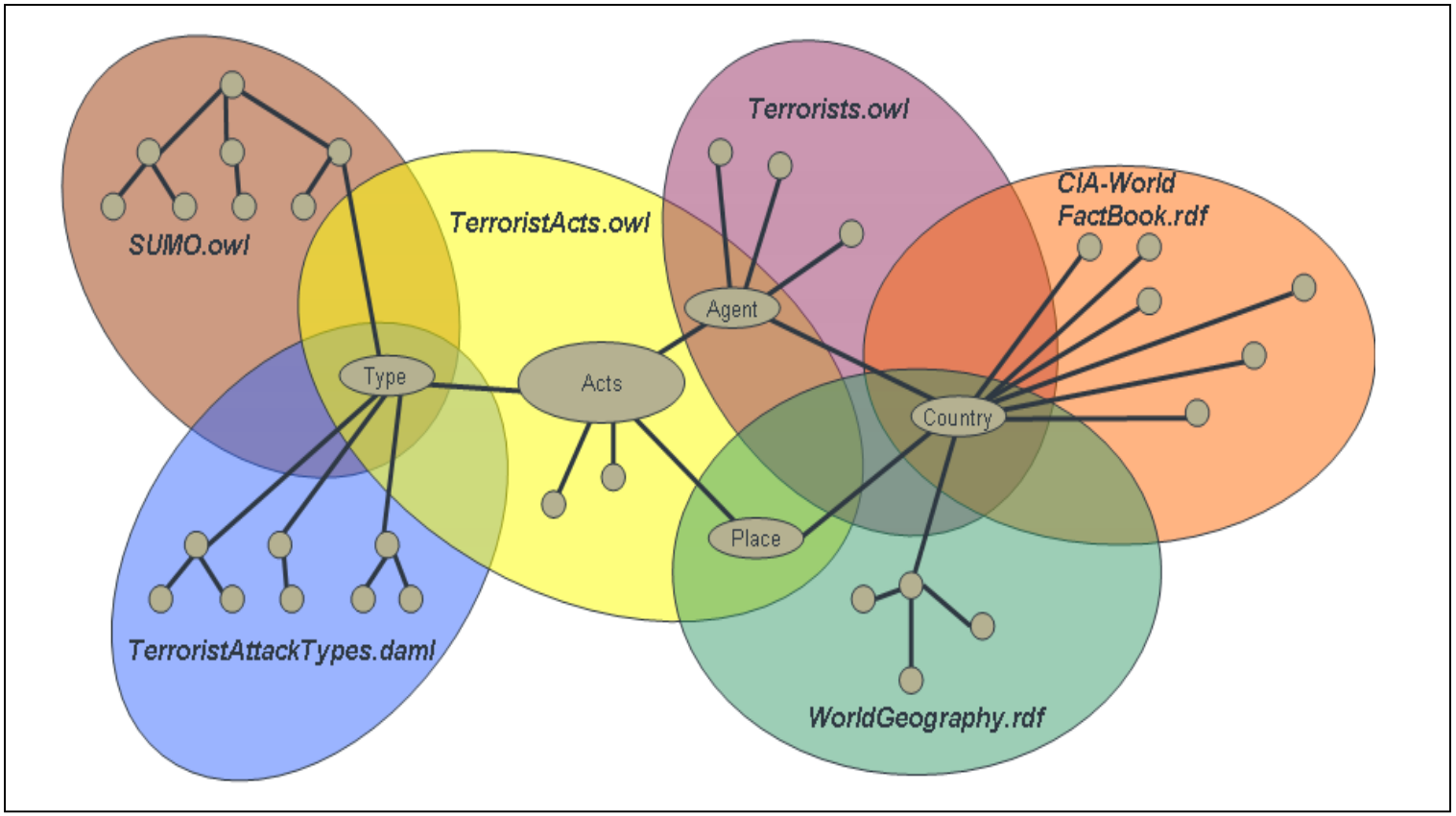
RDF Schema (RDFS)

- 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

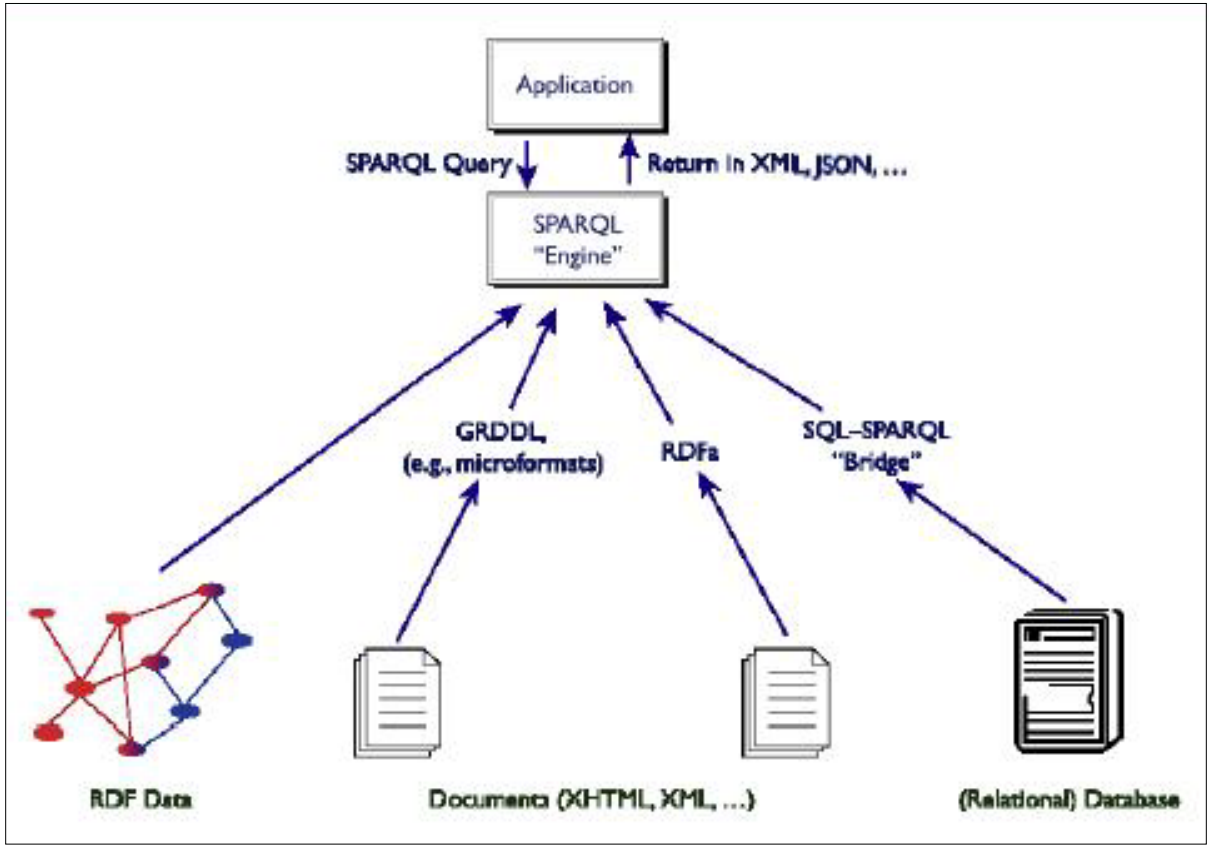


Source: Siderean Software

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

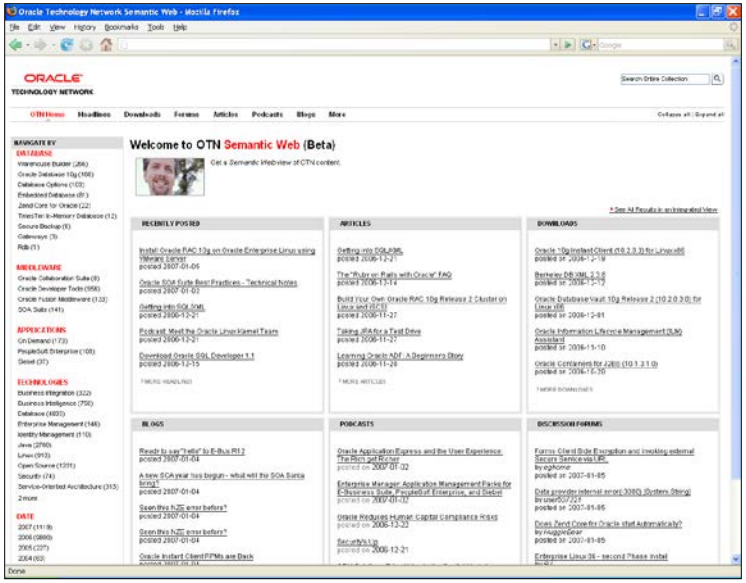
Rules

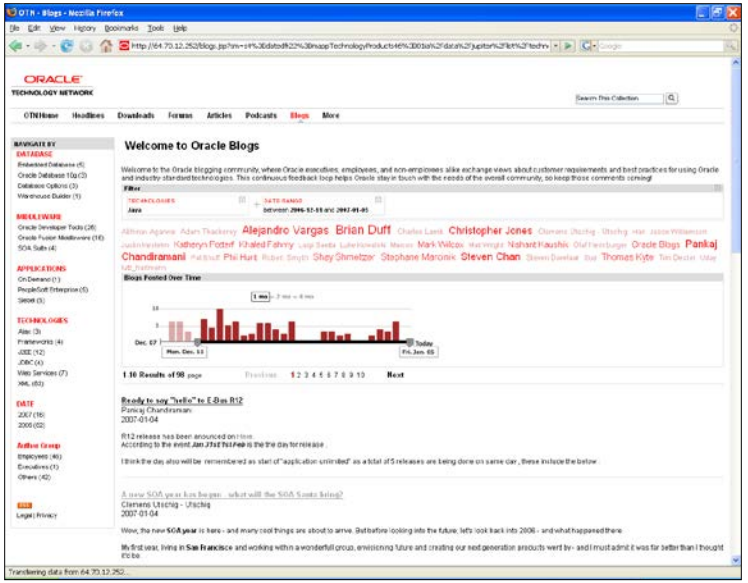
- OWL DL and OWL Lite are based on Description Logic; there are things that DL cannot express
  - Attempts to combine ontologies and rules include RuleML, SWRL, cwm, etc.
- There is also an increasing number of rule-based systems that want to interchange rules
- The W3C Rule Interchange Format Working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



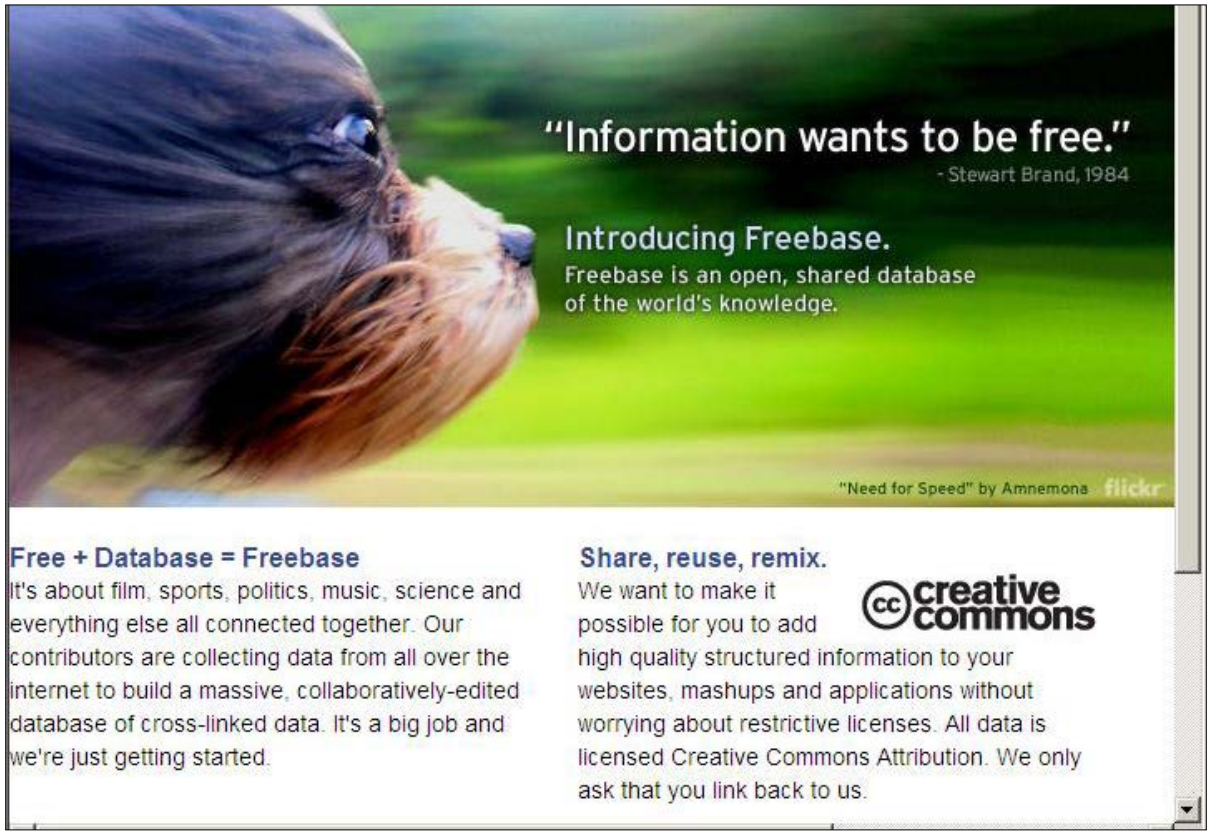


Source: Oracle OTN

Improved Reliability of Search Results

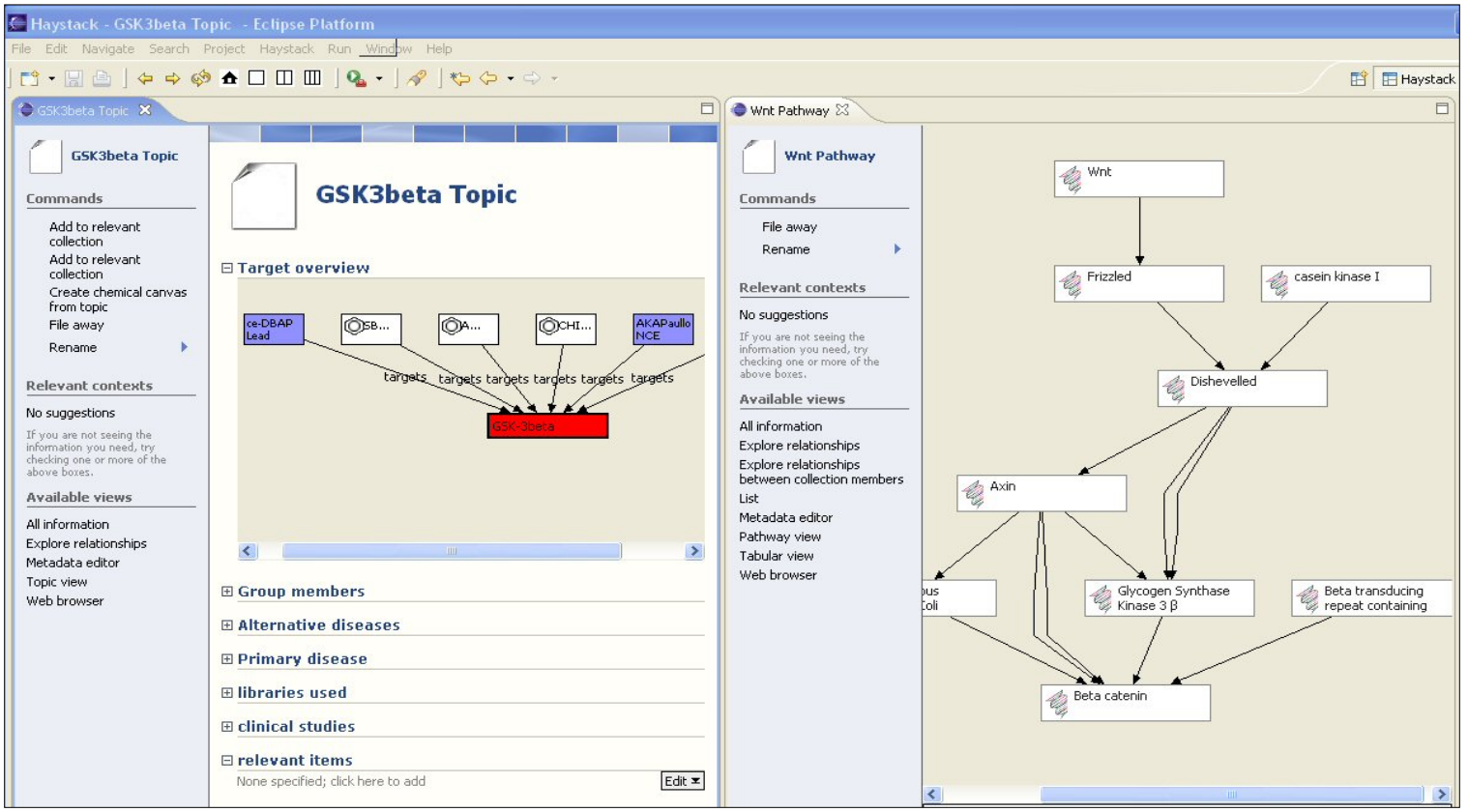
Source: Segala

Web of Data



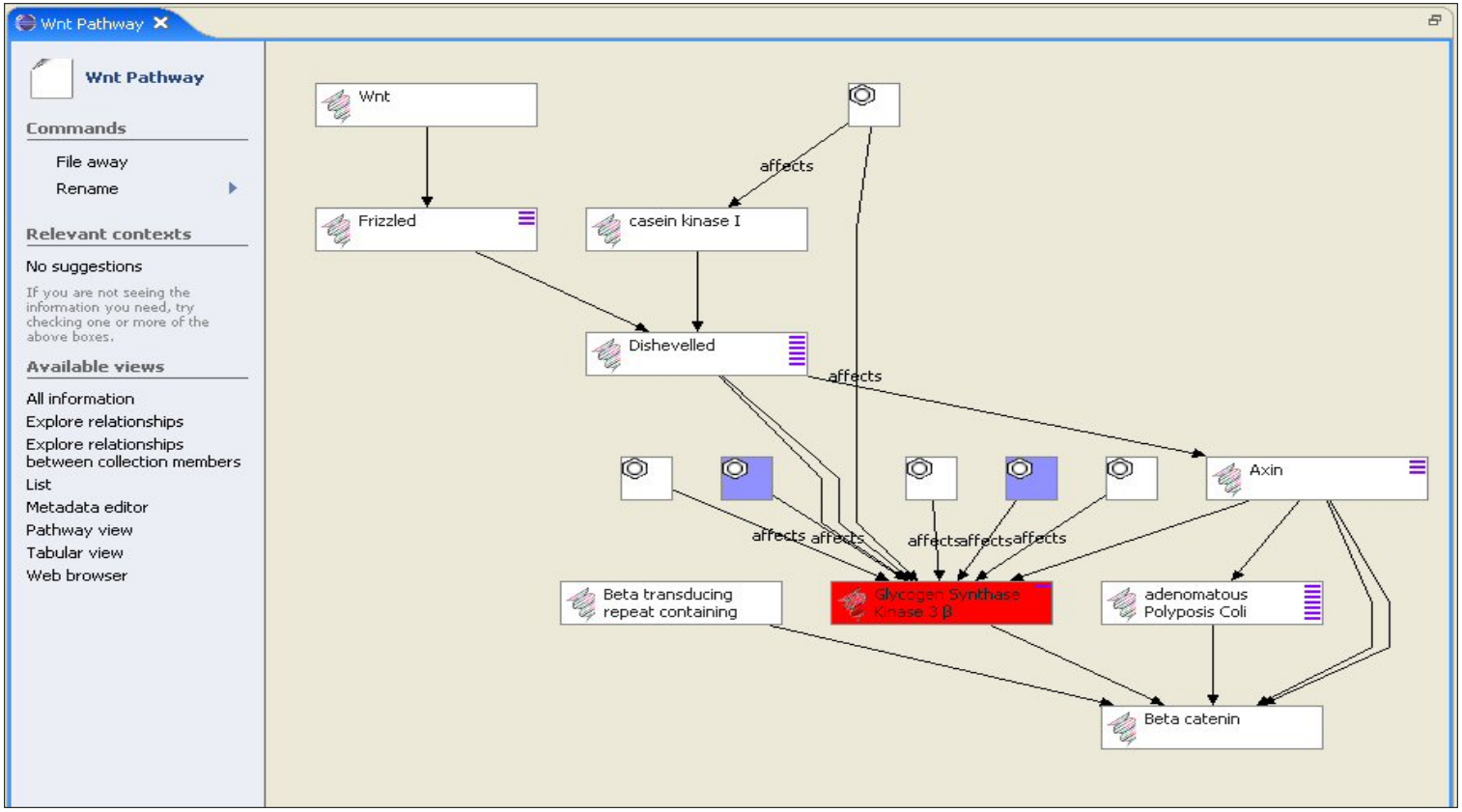
Source: Freebase

Data Mashups



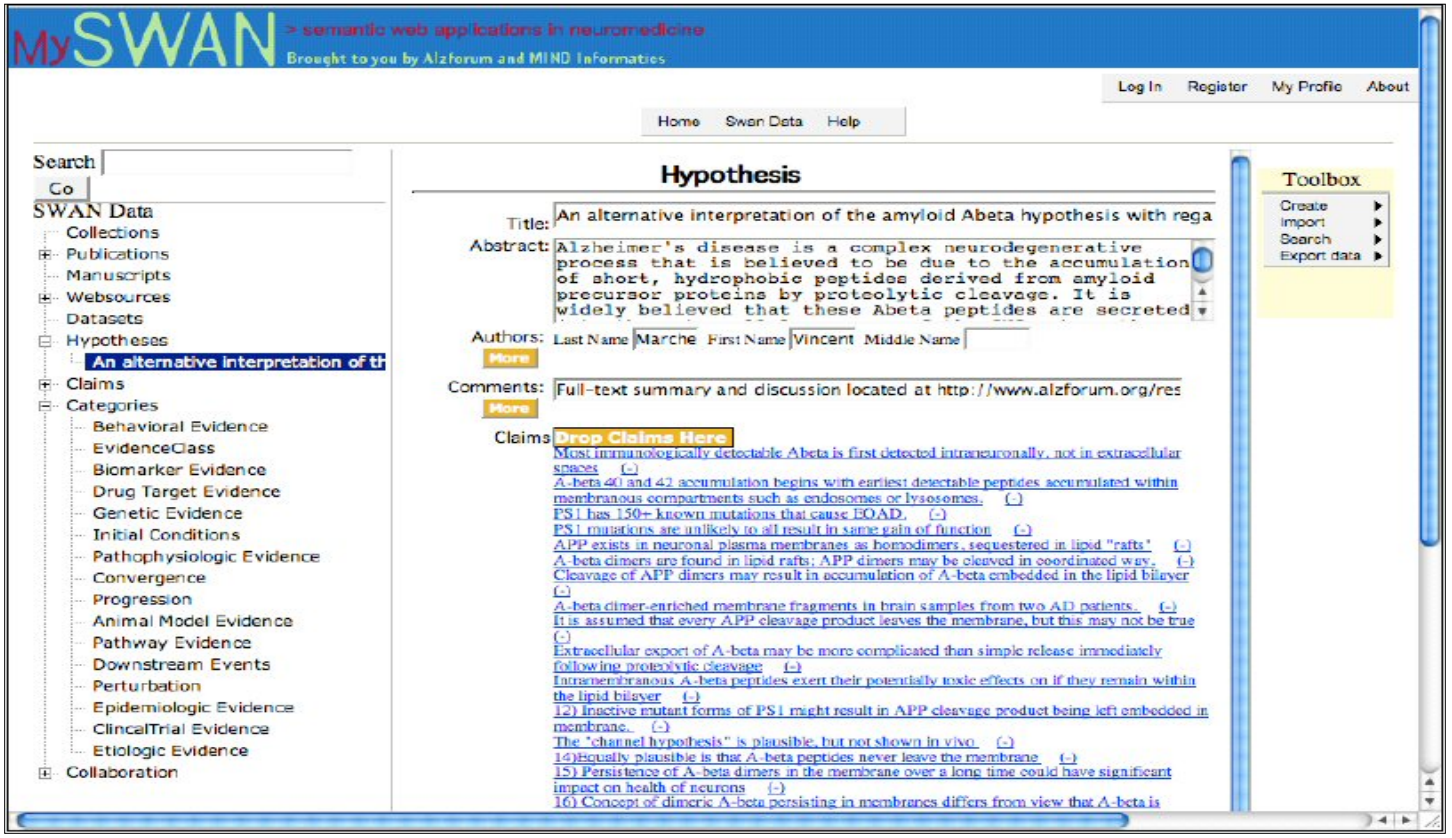
Source: BioDASH

Data Mashups



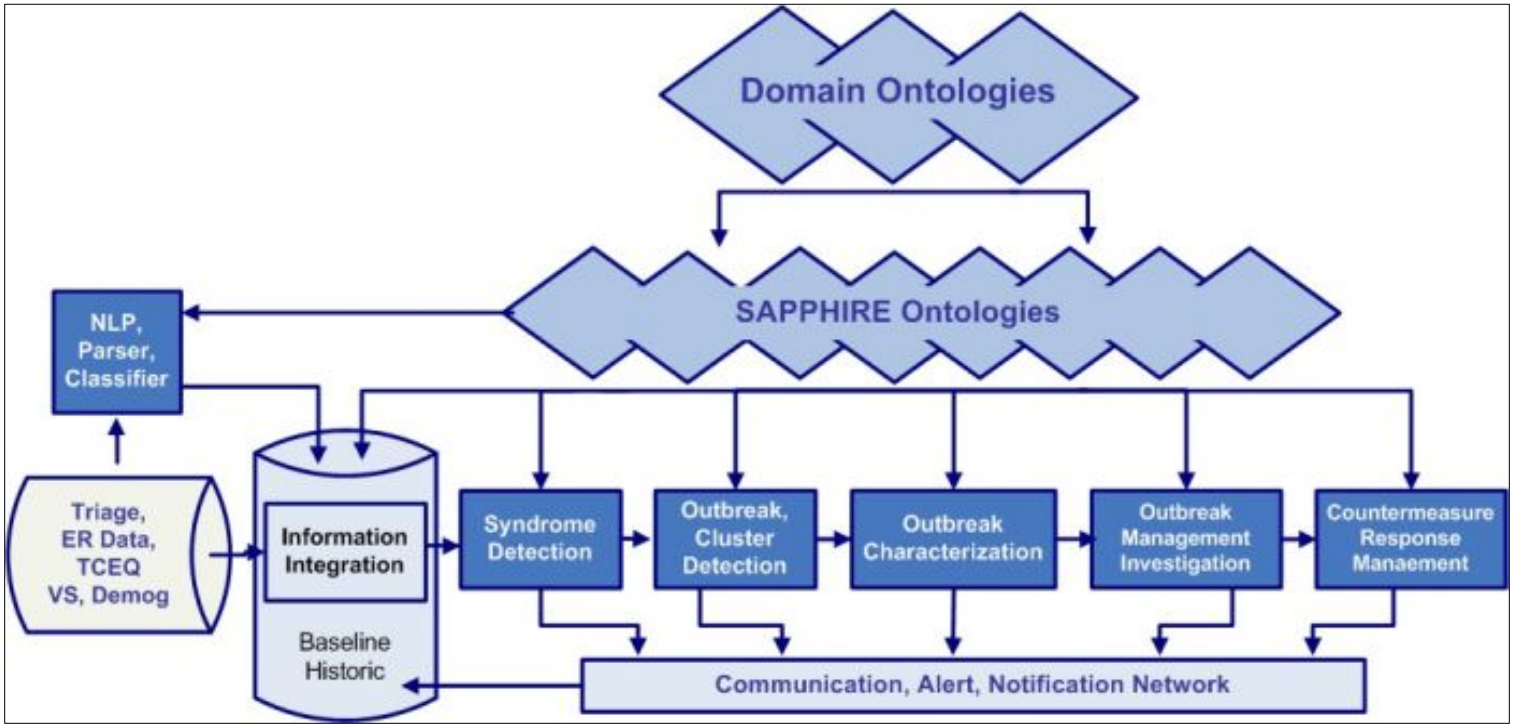
Source: BioDASH

Document Management



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

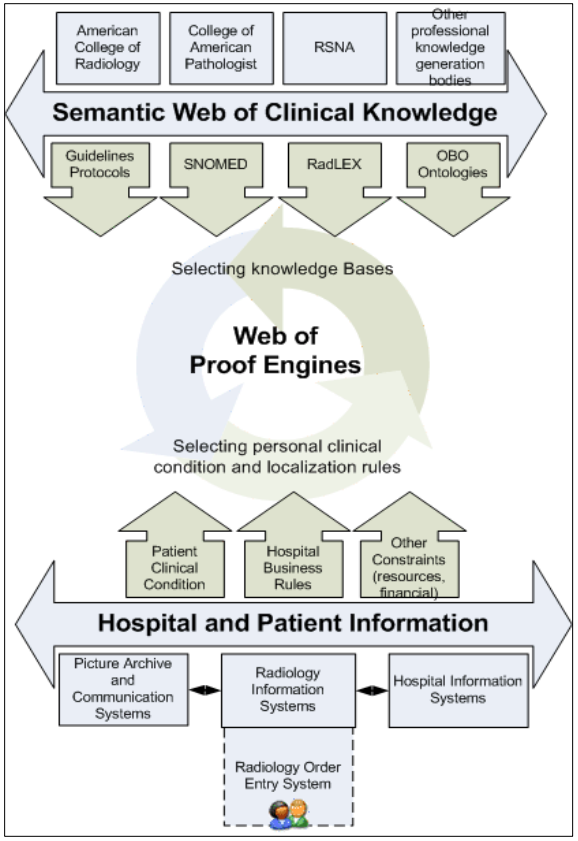
Integration with Semantic Mediation

- 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

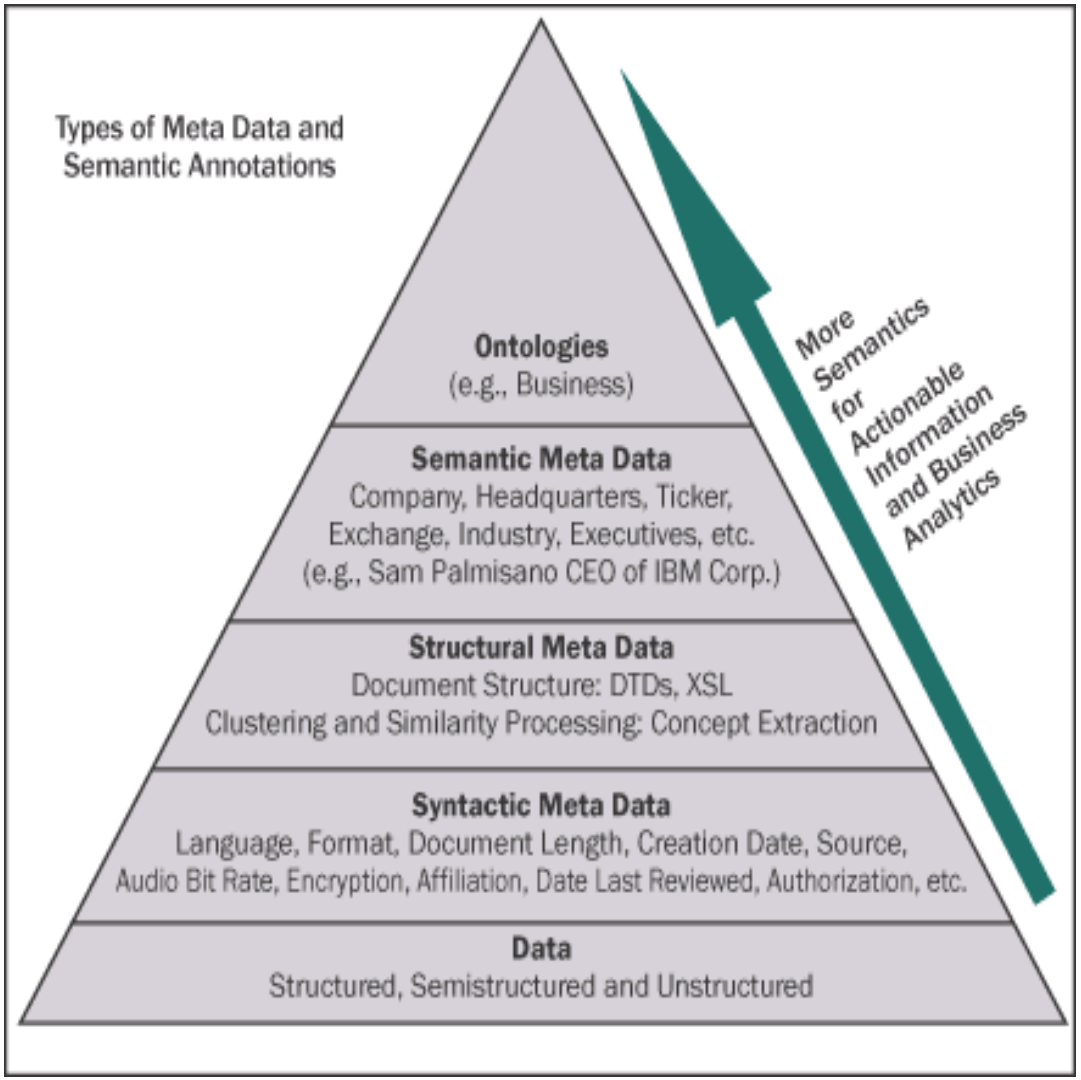
Enablers and Techniques

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 (meatadata Extraction): Manual, Semi-automatic (automatic with human verification), Automatic

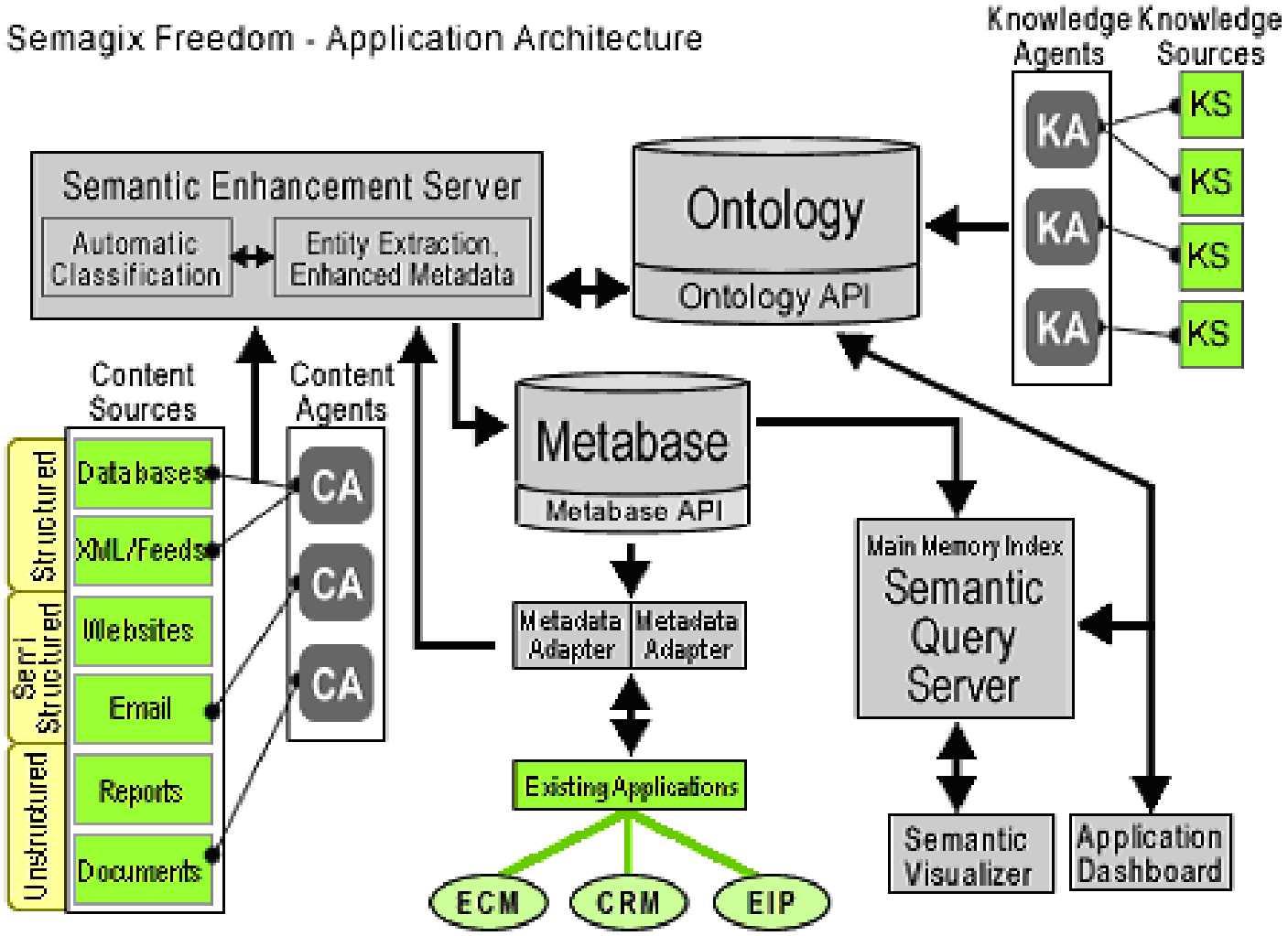
Reasoning/computation: semantics enabled search, integration, complex queries, analysis (paths, subgraph), pattern finding, mining, hypothesis validation, discovery, visualization

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



SEMAGIX
POWER · THROUGH · RELEVANCE

© Semagix, Inc.

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: Descriptional 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

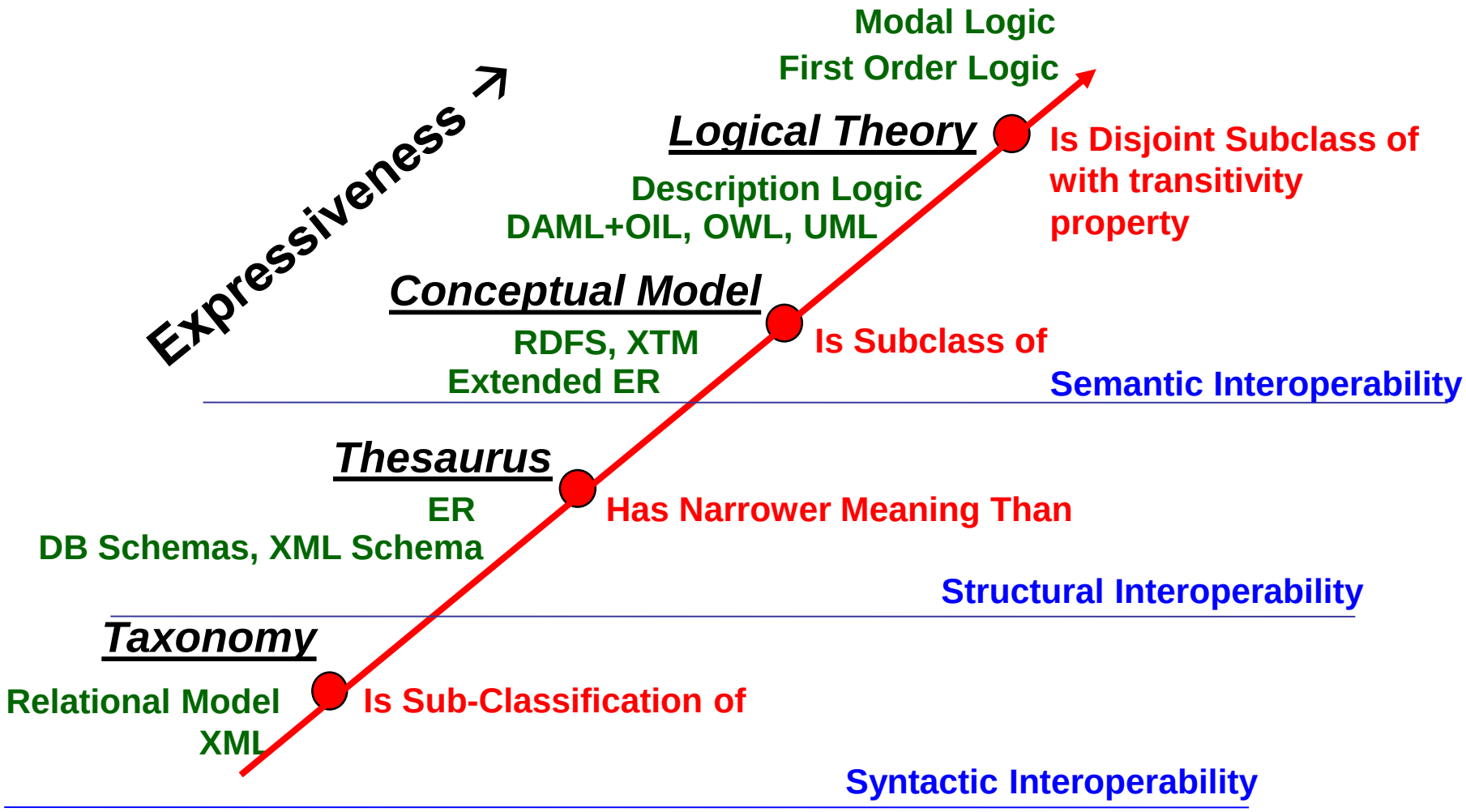
Repertoire Management

Equity Markets

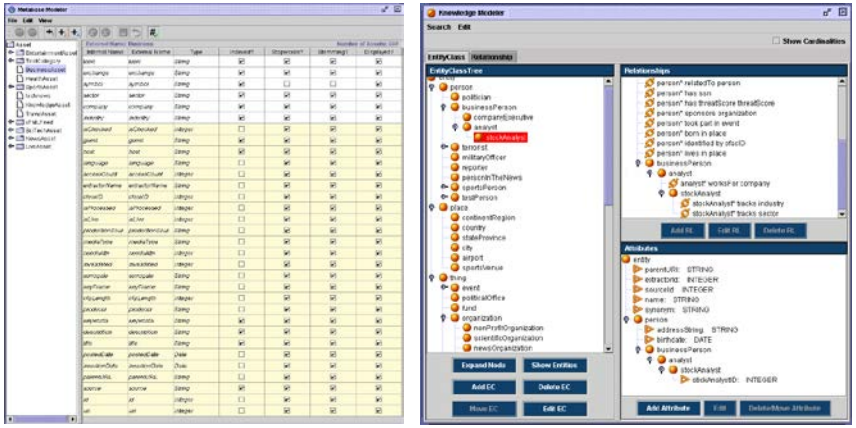
Anti-money Laundering, Financial Risk, Terrorism

Can be Public, Government, Limited Availability, Commercial

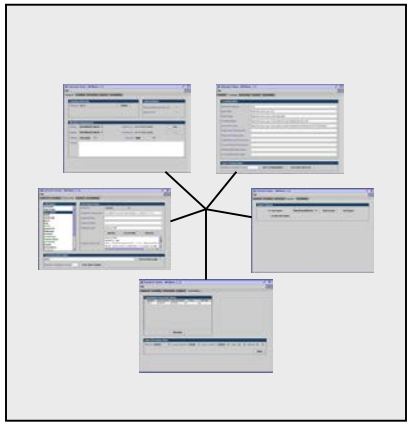
Ontology Language/ Representation Spectrum



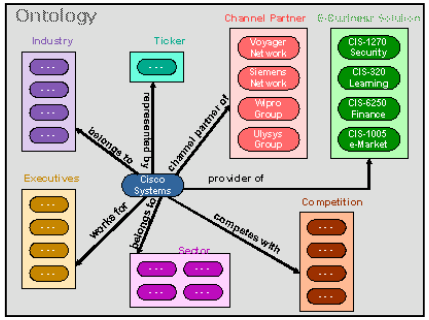
Ontology Creation and Maintenance Steps (Approach 1)



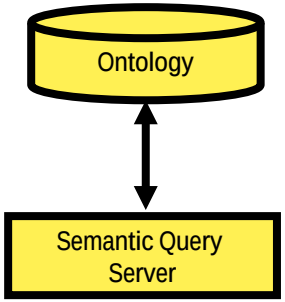
1. Ontology Model Creation (Description)



2. Knowledge Agent Creation



3. Automatic aggregation of Knowledge



4. Querying the Ontology

SEMAGIX
POWER · THROUGH · RELEVANCE

© 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

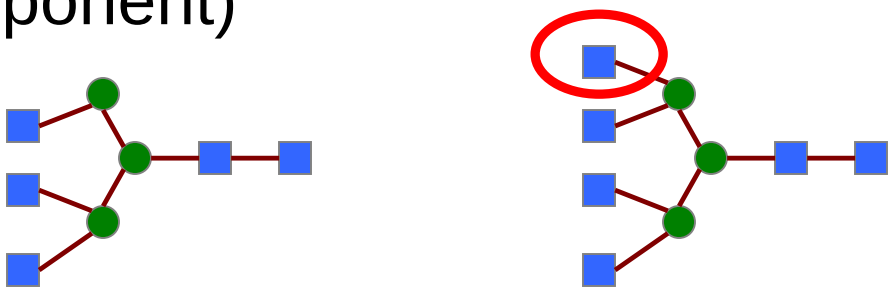
Published through the National Center for Biomedical Ontology (NCBO) and Open Biomedical Ontologies (OBO)

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

- KEGG - Kyoto Encyclopedia of Genes and Genomes
- SWEETDB
- 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

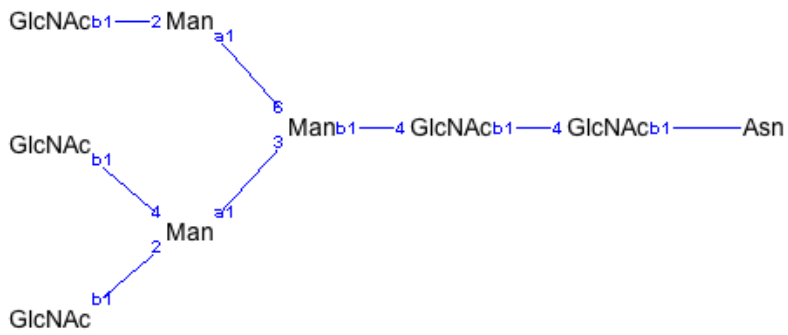
Some sources can be wrong, but not all will be

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

Ontology population workflow

KEGG **GLYCAN: G00020**

Help

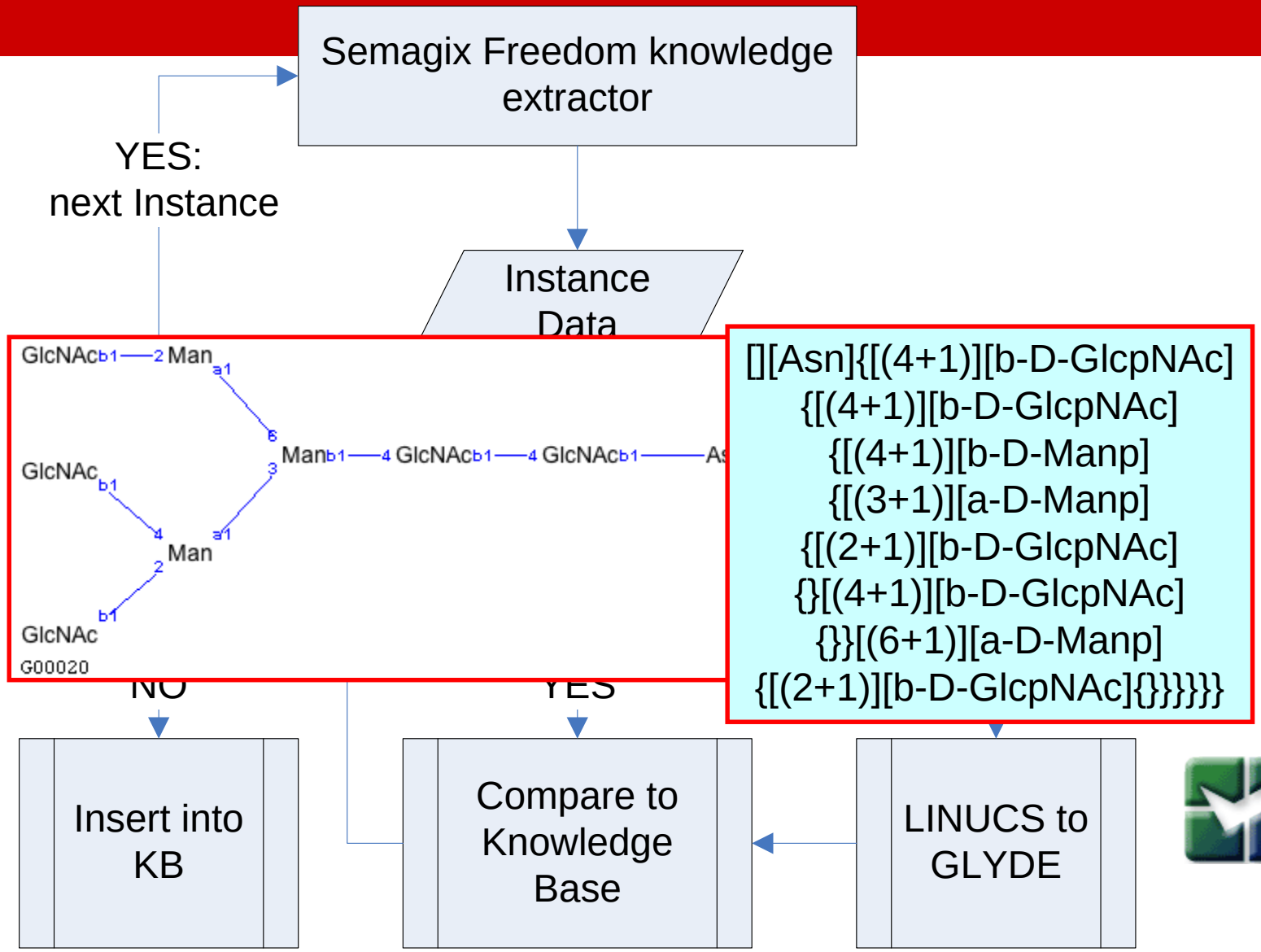
Entry	G00020 Glycan
Composition	(GlcNAc)5 (Man)3 (Asn)1
Mass	1502.4 (Asn)
Structure	 G00020 KCF file KCaM
Class	Glycoprotein; N-Glycan
Reaction	R05987 R05991
Pathway	PATH: map00510 N-Glycan biosynthesis
Enzyme	2.4.1.145 2.4.1.155
Ortholog	KO: K00738 alpha-1,3-mannosylglycoprotein beta-1,4-N-acetylglucosaminyltransferase KO: K00744 alpha-1,3(6)-mannosylglycoprotein beta-1,6-N-acetyl-glucosaminyltransferase
Other DBs	CCSD: 4362 5058 7636 30028 33048 34937 35723 42024 43245
LinkDB	All DBs
KCF data	Show

C to
JCS

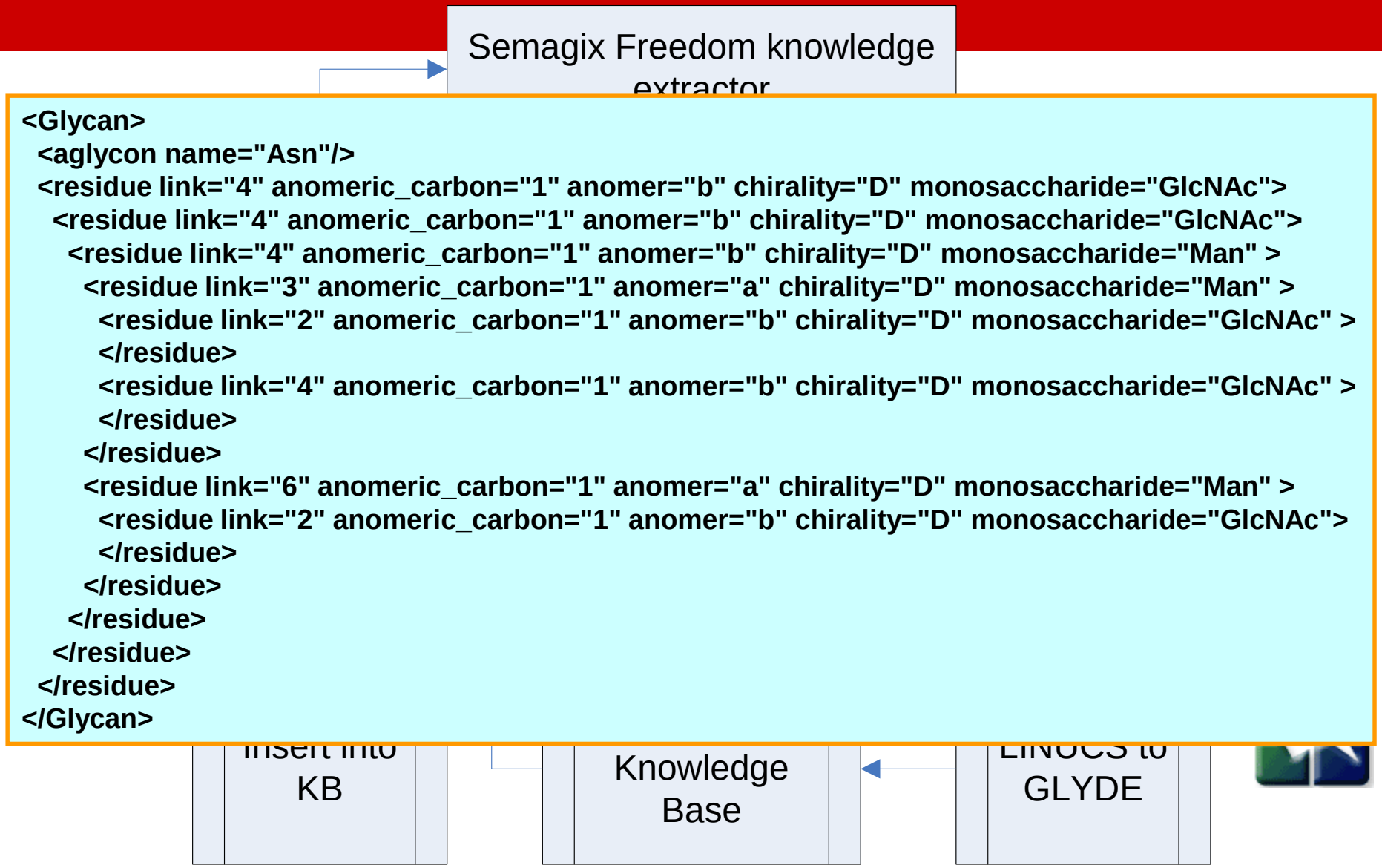
CS to
DE

Base

Ontology population workflow

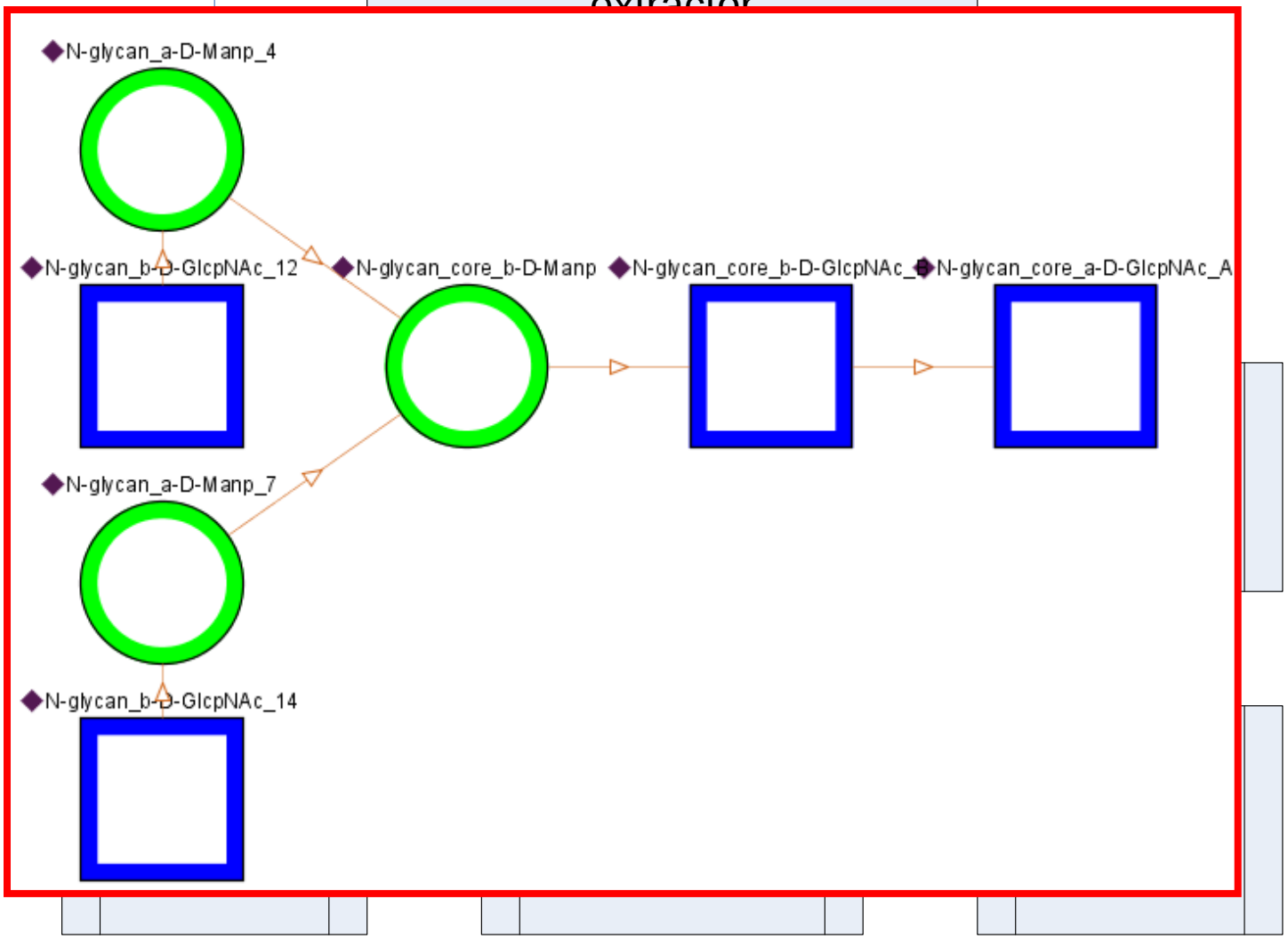


Ontology population workflow



Ontology population workflow

Semagix Freedom knowledge extractor



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

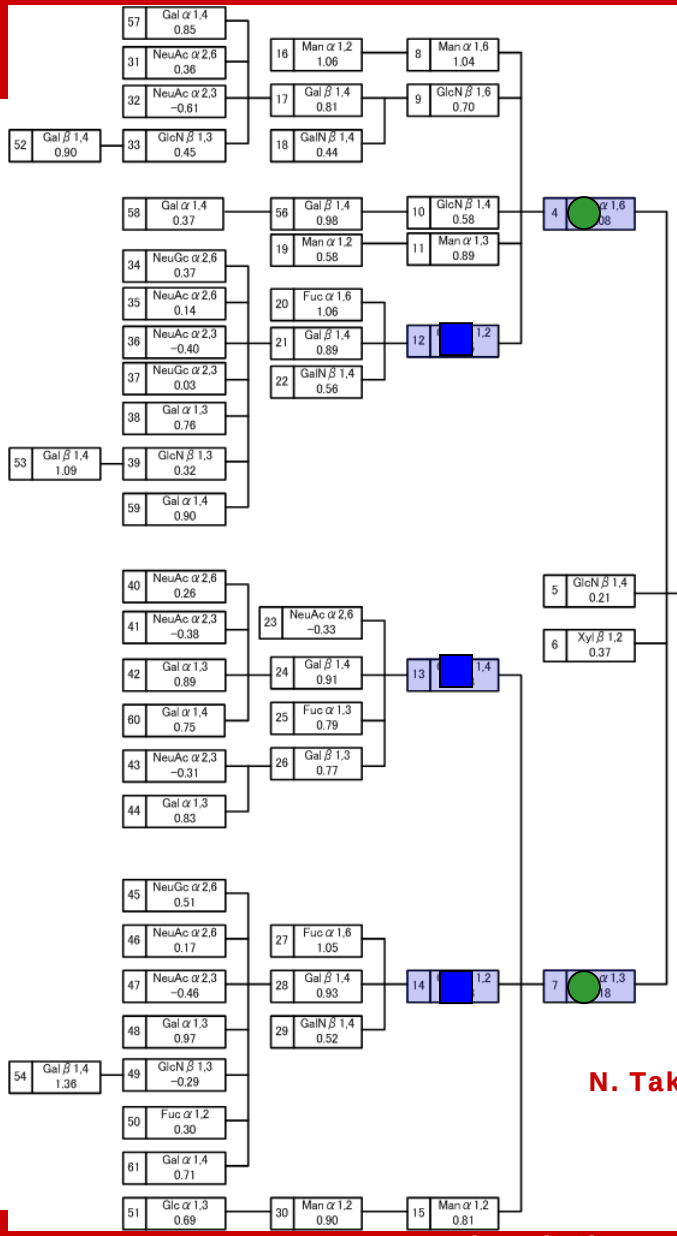
?

Errors in KEGG and Carbbank are the same

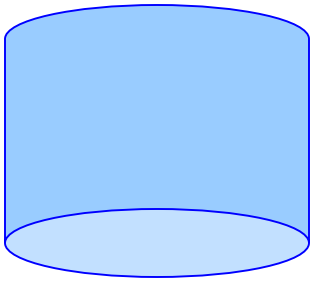
Cannot use these sources for comparison

Needs curation by the expert community

GlycoTree

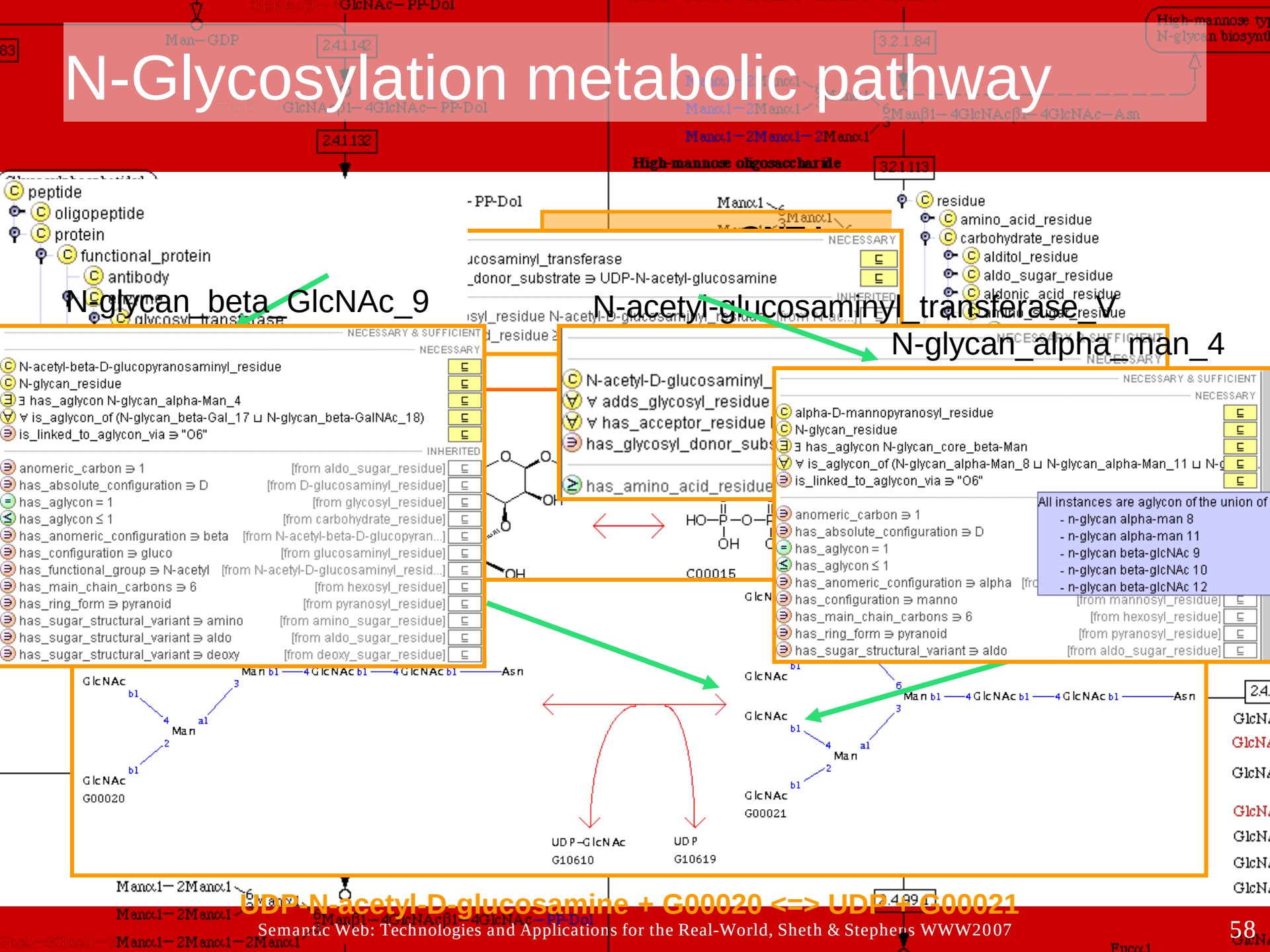


β-D-Glc*p*NAc-(1-2)-α-D-Man*p*-(1-6)+
β-D-Man*p*-(1-4)-β-D-Glc*p*NAc-(1-4)-β-D-Glc*p*NAc
β-D-Glc*p*NAc-(1-4)-α-D-Man*p*-(1-3)+
β-D-Glc*p*NAc-(1-2)+



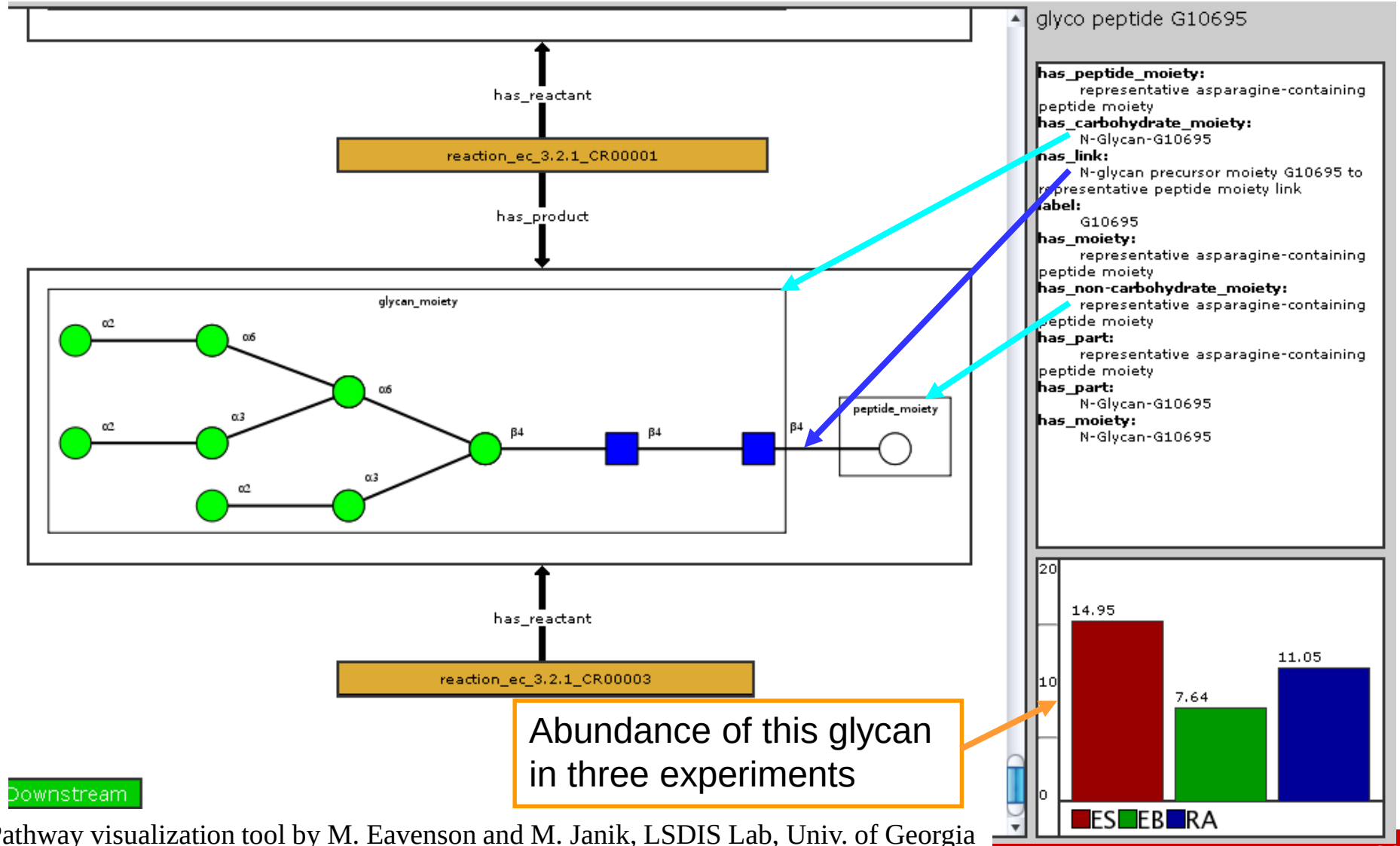
N. Takahashi and K. Kato, *Trends in Glycosciences and Glycotechnology*, 15: 235-251

N-Glycosylation metabolic pathway

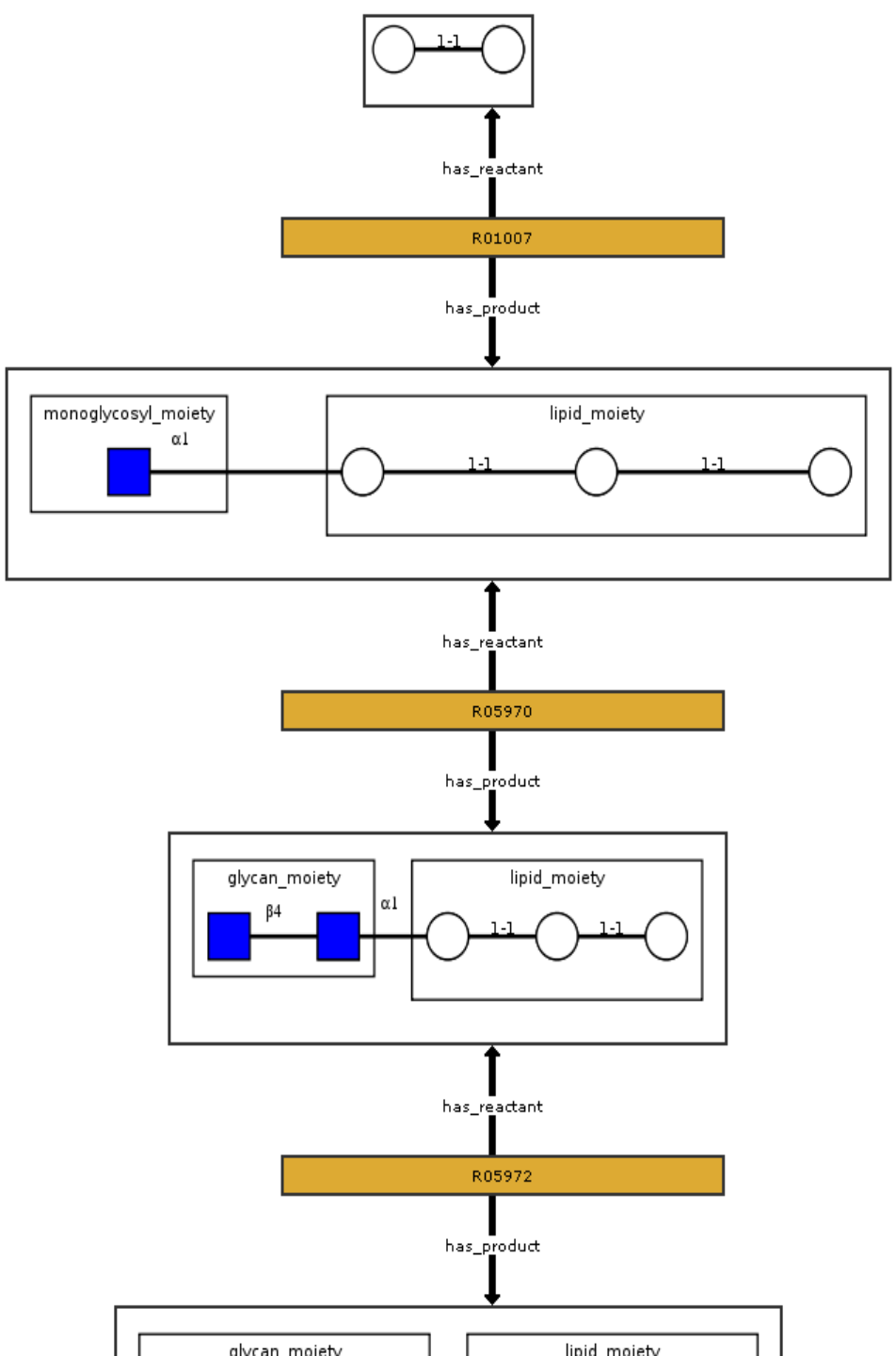


UDP-N-acetyl-D-glucosamine + G00020 <=> UDP + G00021

Pathway Steps - Glycan



Pathway visualization tool by M. Eavenson and M. Janik, LSDIS Lab, Univ. of Georgia


1-1
has_reactant
R01007
has_product
monoglycosyl_moiety
α1
lipid_moiety
1-1
1-1
has_reactant
R05970
has_product
glycan_moiety
β4
α1
lipid_moiety
1-1
1-1
has_reactant
R05972
has_product

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 fore is active on the southern perimeter, which i[illegible] black spruce. The
fire has increased in size, but was not ma[illegible]ter is
35% contained, while protection of the [illegible]

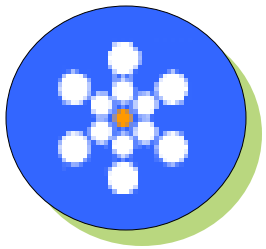
Date => day month int ',' int

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.

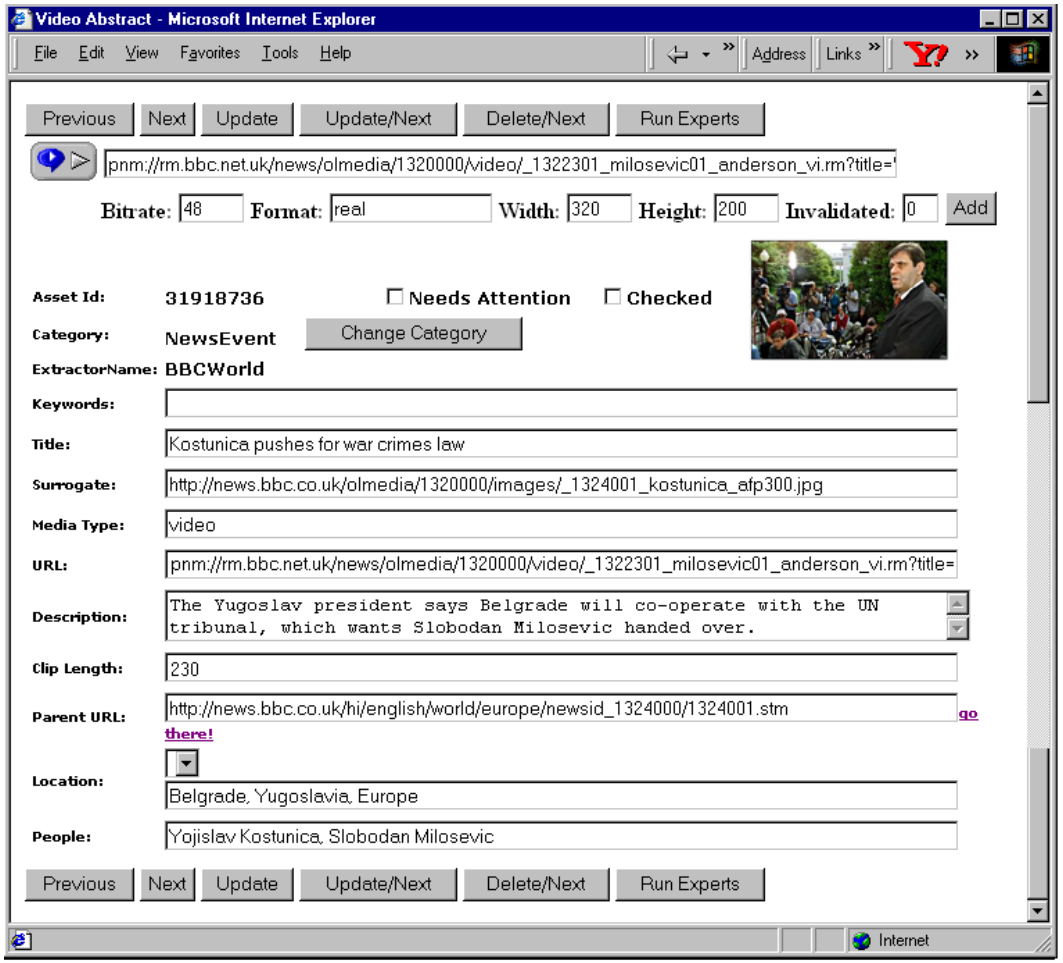
Taalee Extraction and Knowledgebase Enhancement

Web Page

Extraction Agent

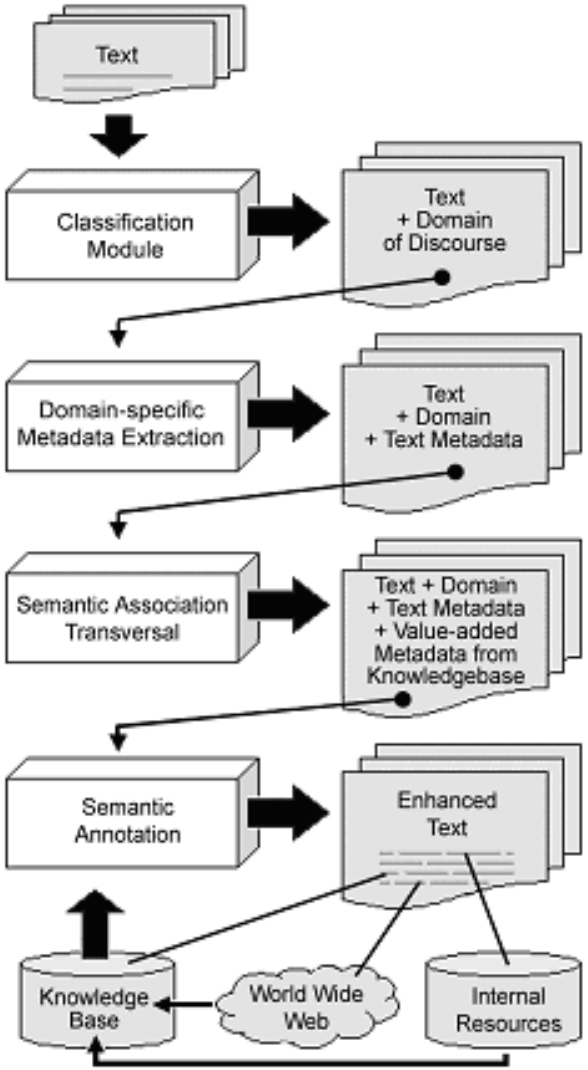


Enhanced Metadata Asset



Taalee Semantic Engine, also IBM Semantics Tools

Metadata Extraction and Semantic Enhancement



Semantic Enhancement Engine
[Hammond, Sheth, Kochut 2002]
Also, WebFountatin,
KIM from OntoText

SEMAGIX
POWER · THROUGH · RELEVANCE

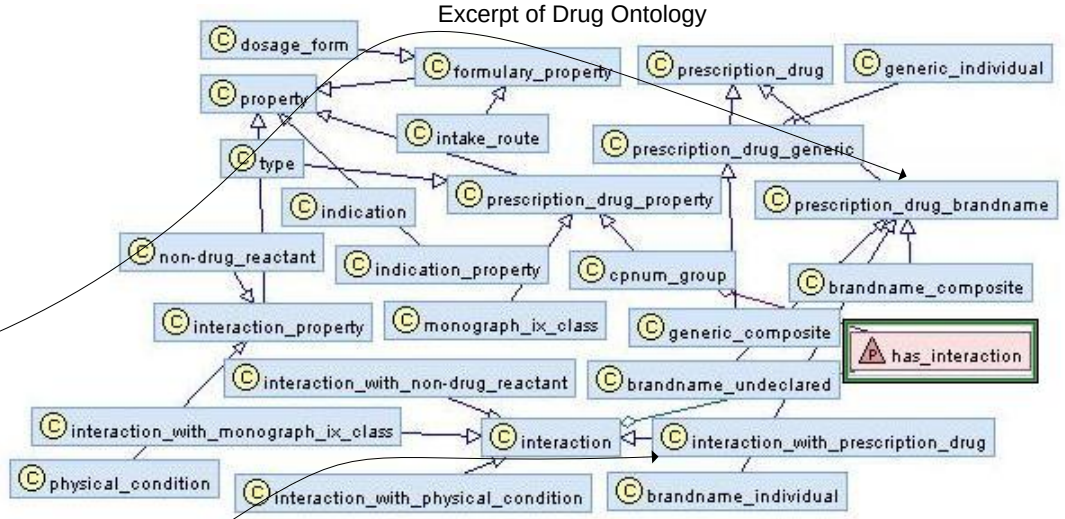
Semantic Annotation - document

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 annou
<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: <Regexp type="date">April 7, 2005</Regexp>

Excerpt of Drug Ontology

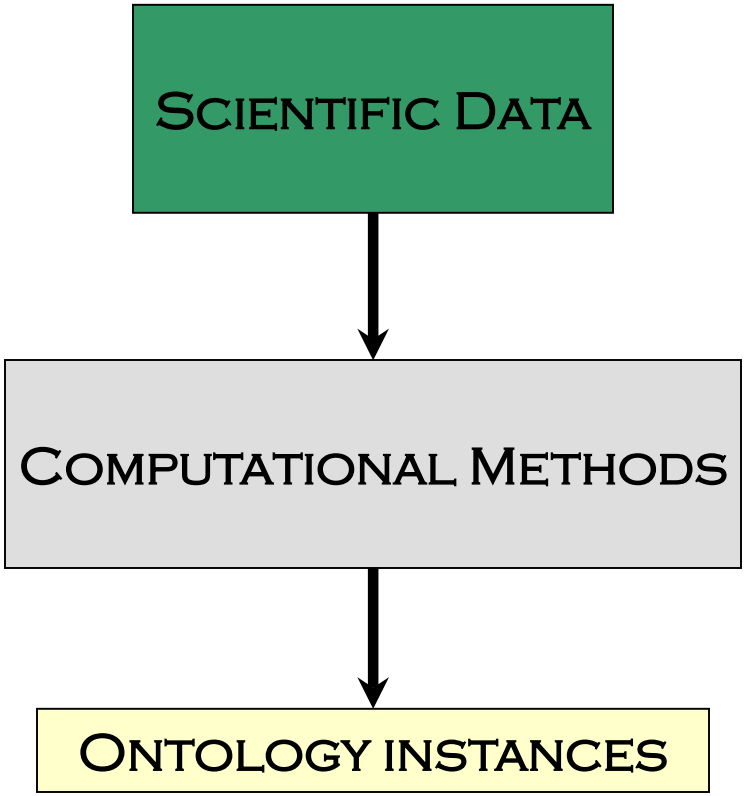


Sample Created Metadata

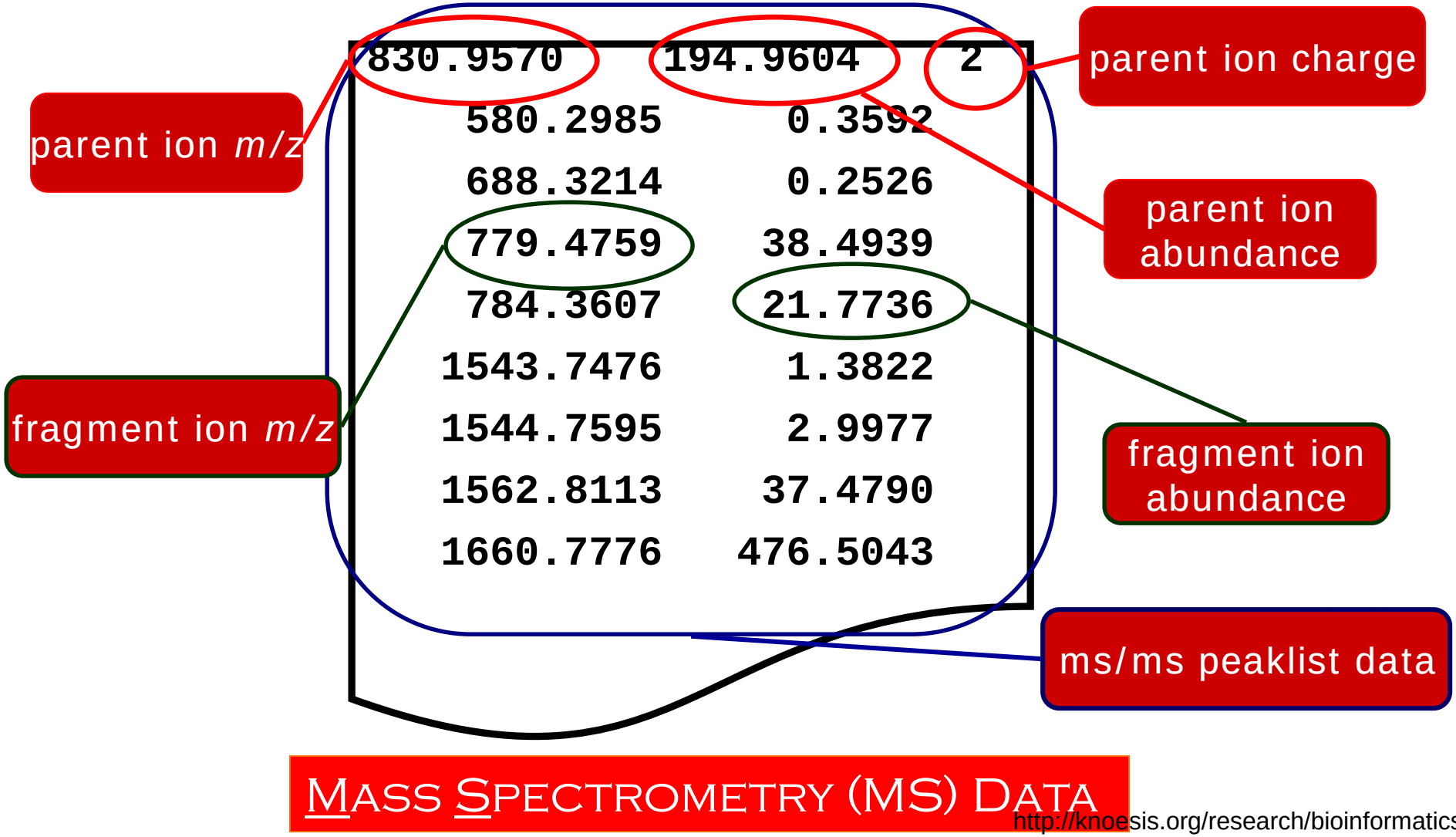
```
<Entity id="122805"
class="DrugOntology#prescription_drug_brandname">
Bextra
  <Relationship id="442134"
    class="DrugOntology#has_interaction">
    <Entity id="14280" class="DrugOntology
          #interaction_with_physical_condition>sulfa allergy
    </Entity>
  </Relationship>
</Entity>
```

Semantic annotation of scientific/experimental data

ProPreO population: transformation to rdf

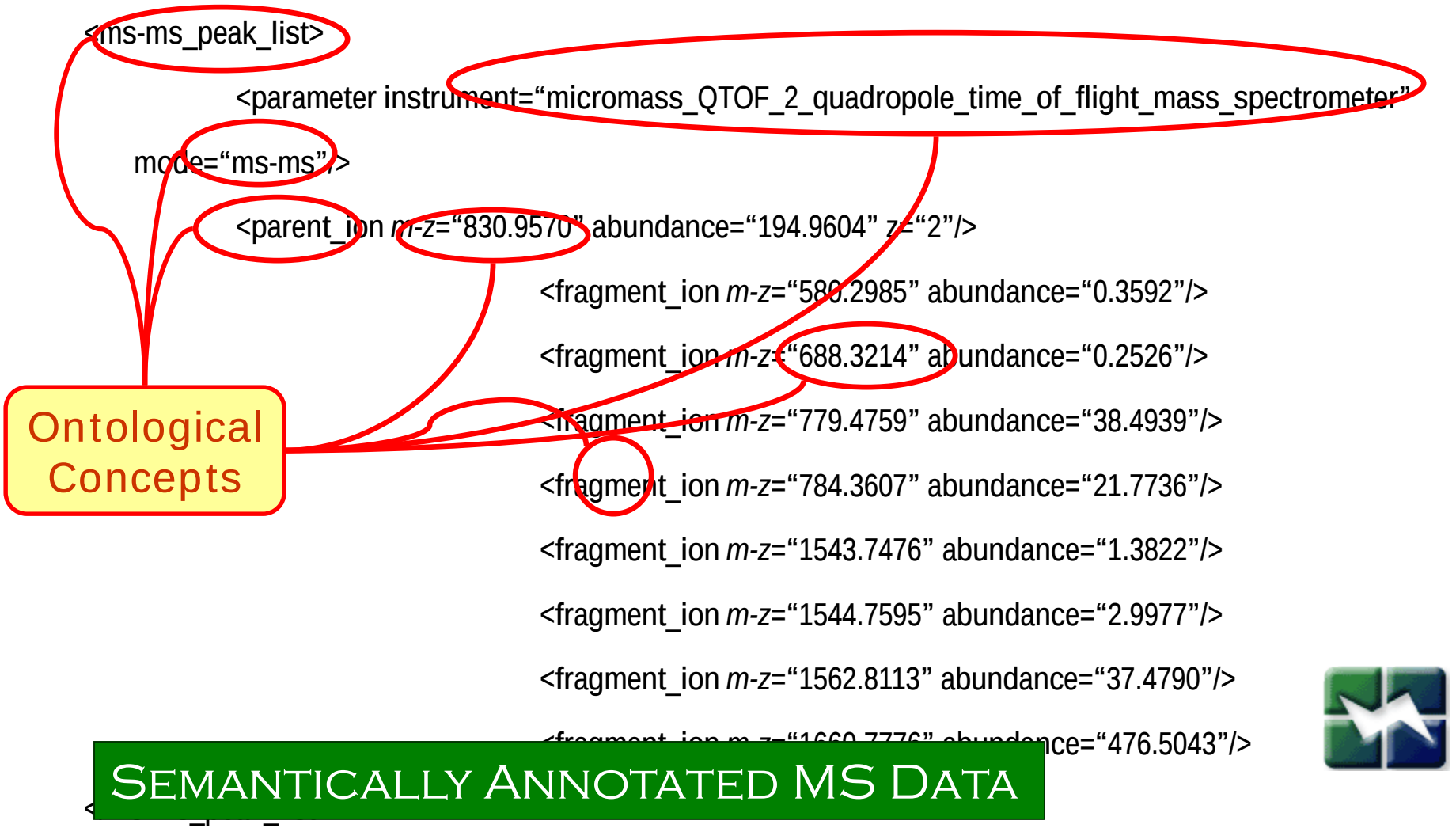


ProPreO: Ontology-mediated provenance



http://knoesis.org/research/bioinformatics

ProPreO: Ontology-mediated provenance



http://knoesis.org/research/bioinformatics

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

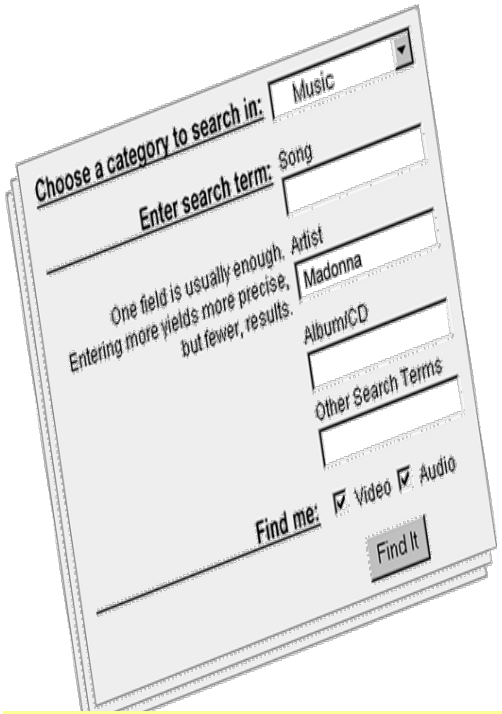
Standards Usage

Scalability

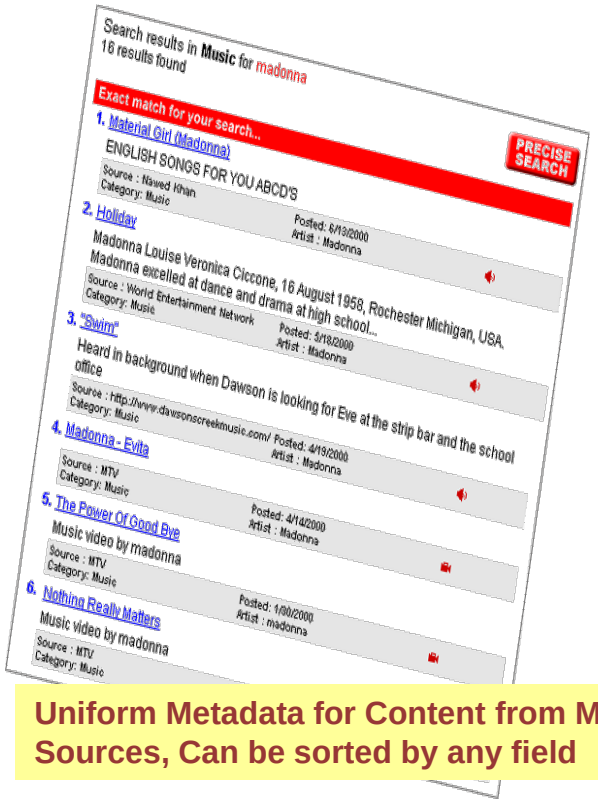
Performance

Taalee's Semantic Search

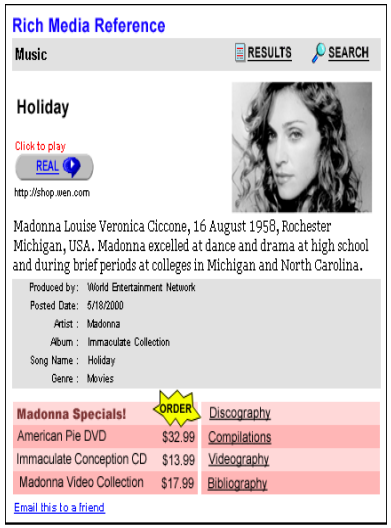
Highly customizable, precise and freshest A/V search



Context and Domain Specific Attributes



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



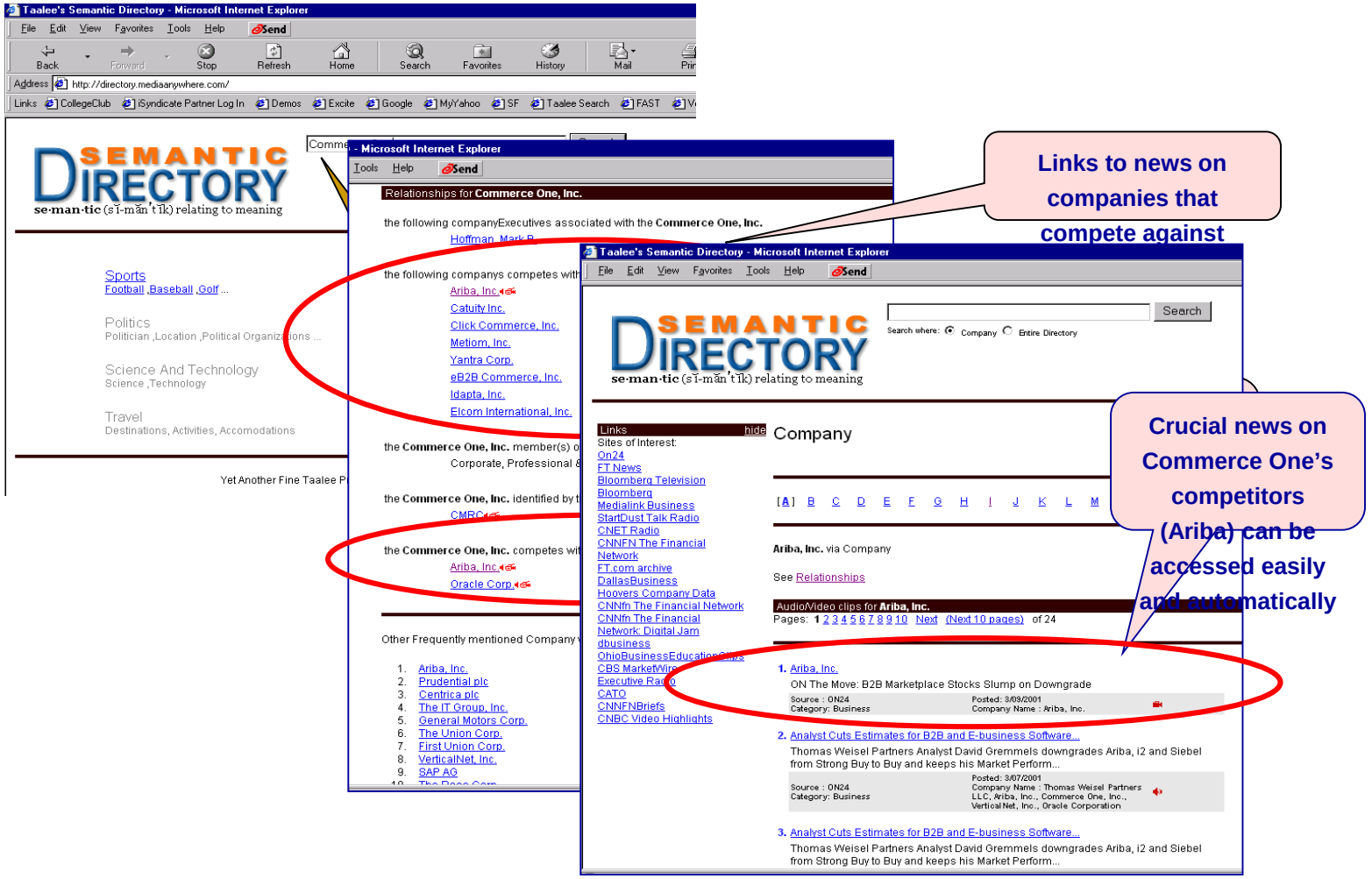
Delightful, relevant information, exceptional targeting opportunity

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

What can a context do?

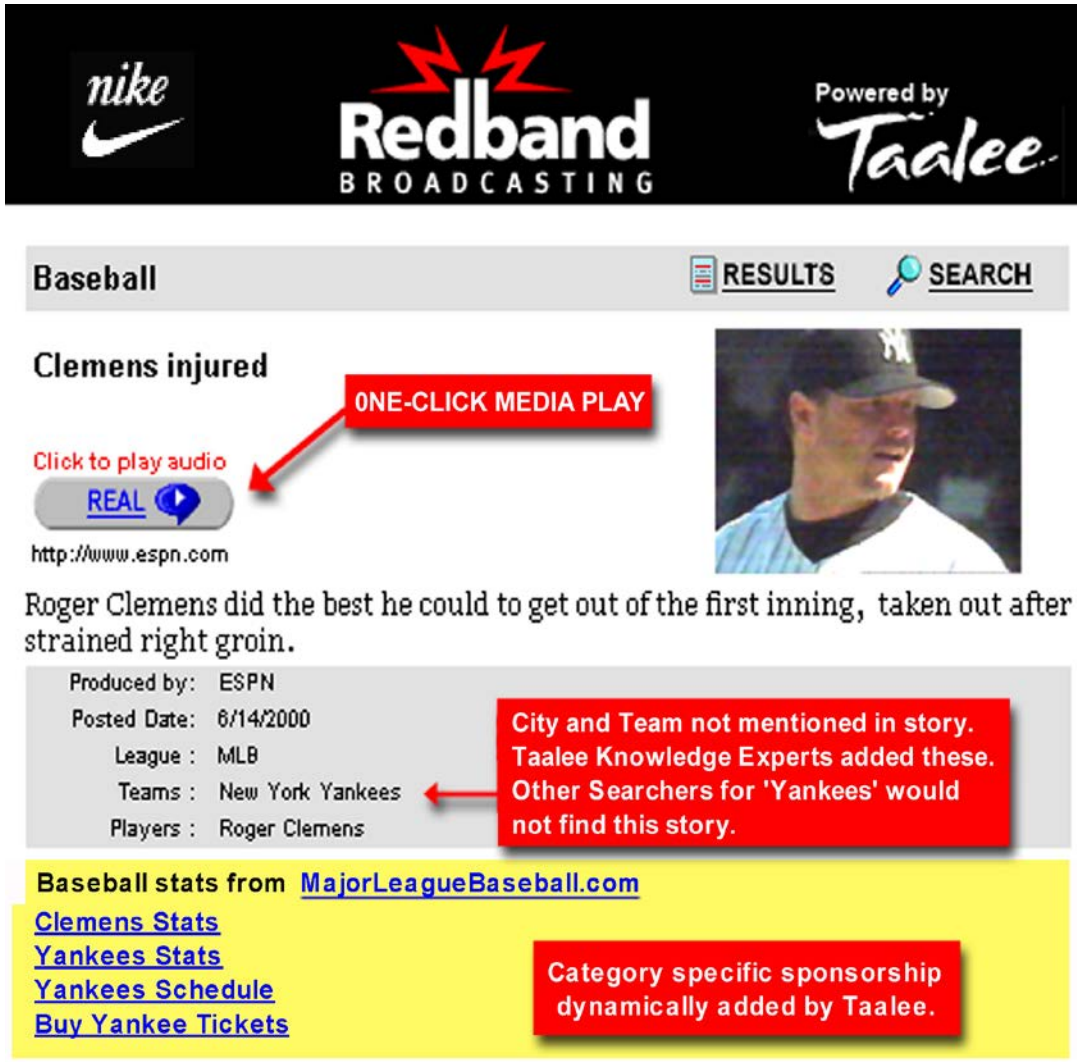
Creating a Web of related information

Ontology driven Semantic Directory



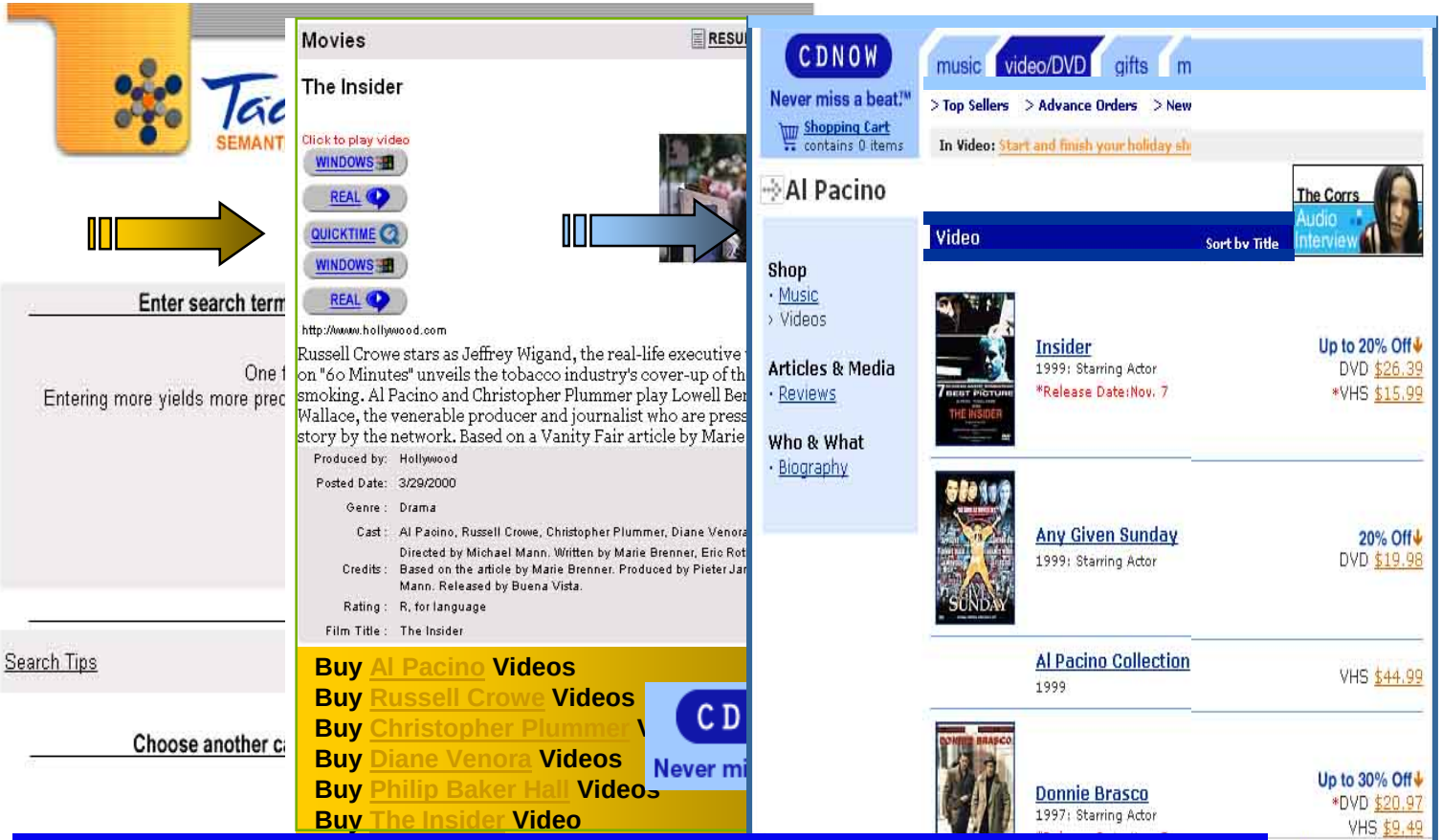
Semantic Targeting

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



Semantic Enrichment

Semantic/Interactive Targeting



Precisely targeted through the use of Structured Metadata and integration from multiple sources

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

Managing Semantic Content on the Web, Internet Computing 2002

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

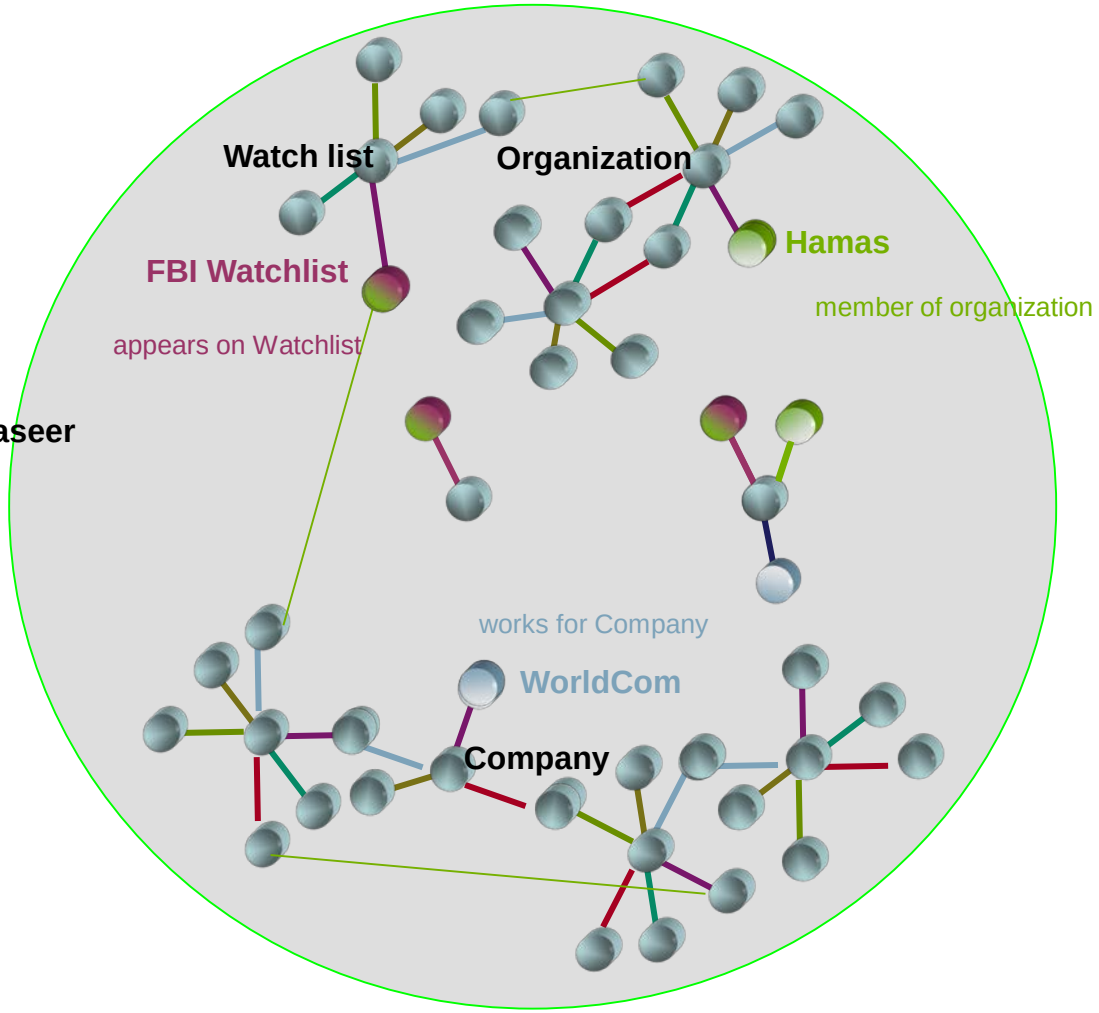
Solution

Rapid and accurate KYC checks

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

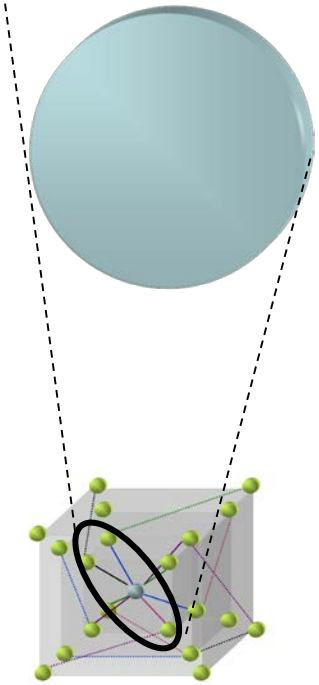
The Process

Ahmed Yaseer



Ahmed Yaseer:

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



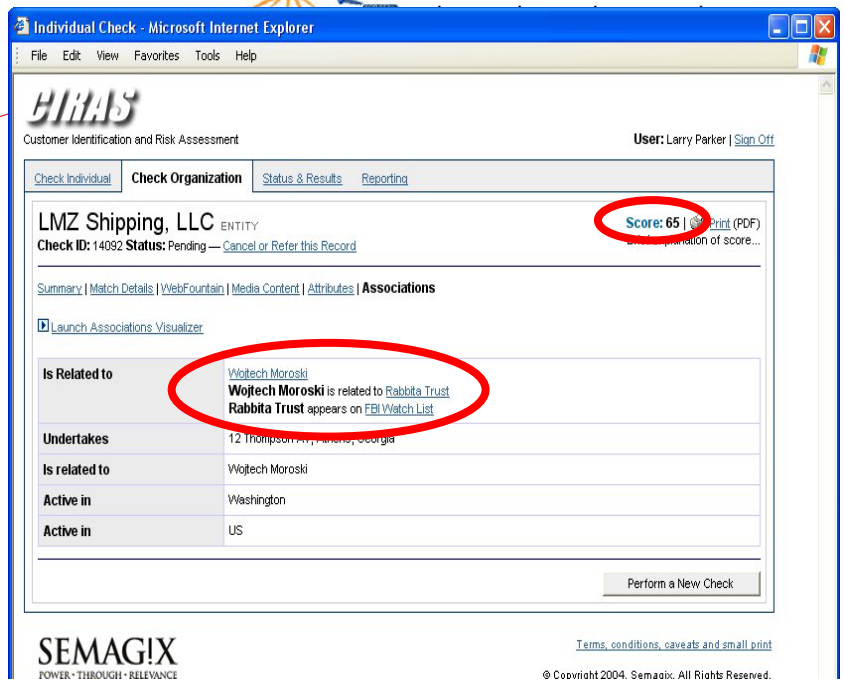
Global Investment Bank

Semi-structured Government Data

Un-structure text, Semi-structured Data

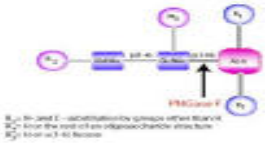
Establishing New Account

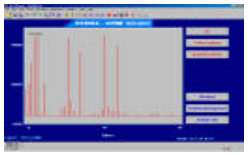
User will be able to navigate

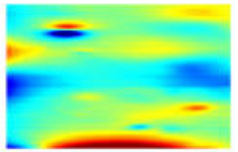


Example of Fraud Prevention application used in financial services

N-Glycosylation Process (NGP)







Cell Culture

extract

Glycoprotein Fraction

proteolysis

Glycopeptides Fraction

1 — n

Separation technique I

Glycopeptides Fraction

n

PNGase

Peptide Fraction

n — n*m

Separation technique II

Peptide Fraction

Mass spectrometry

ms data

Data reduction

ms peaklist

binning

N-dimensional array

Data correlation

ms/ms data

Data reduction

ms/ms peaklist

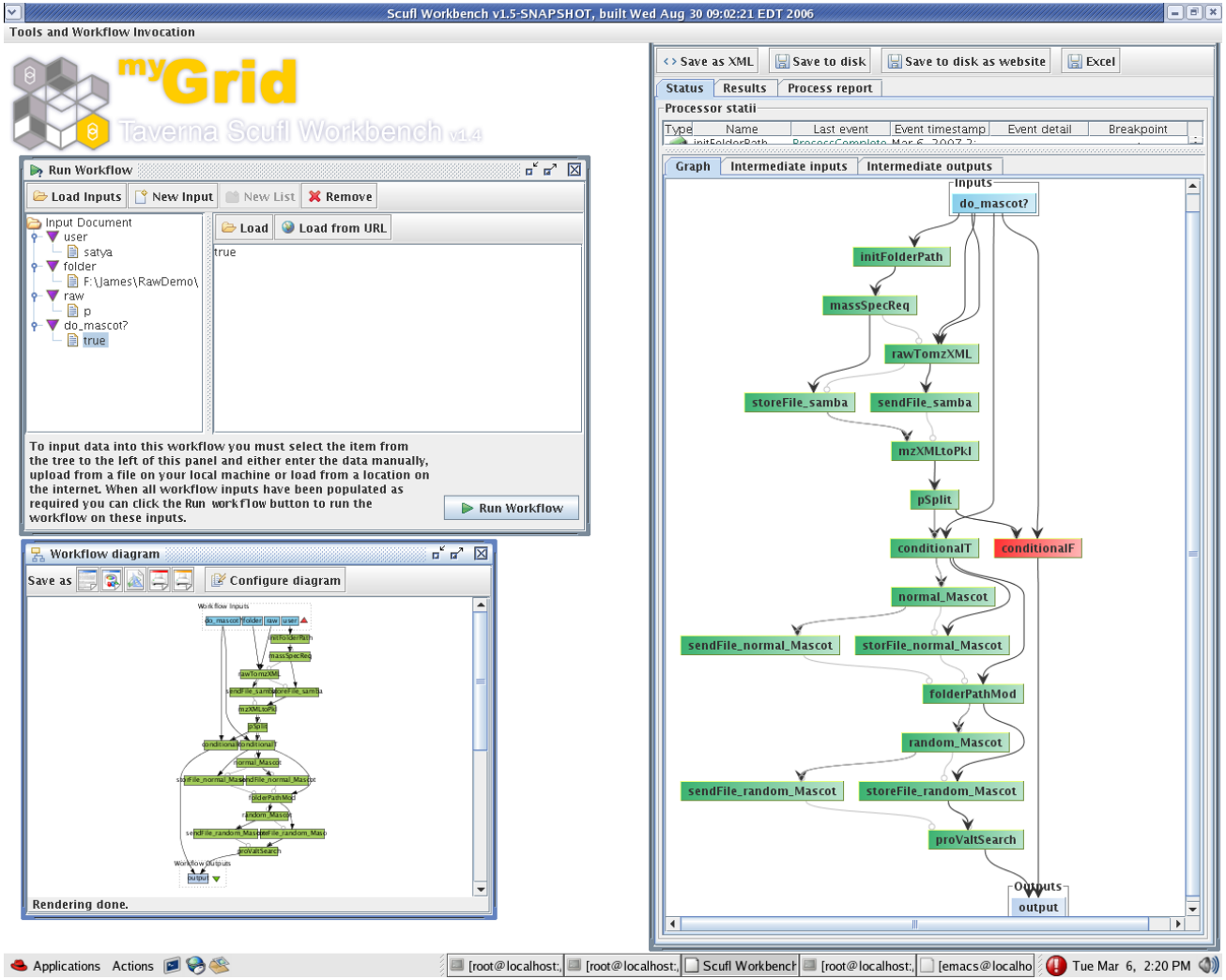
Peptide identification

Peptide list

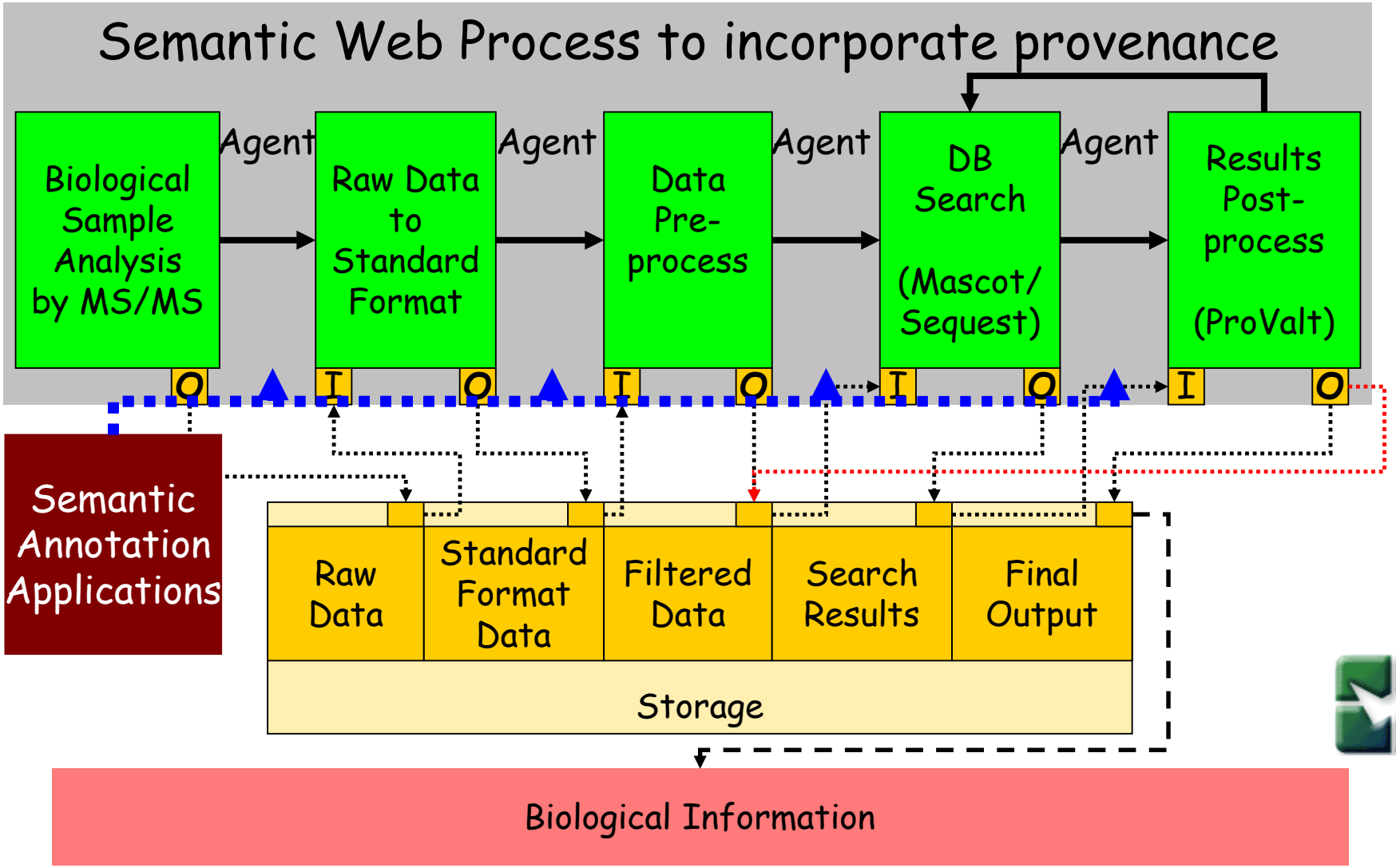
Glycopeptide identification and quantification

Signal integration

Workflow based on Web Services = Web Process

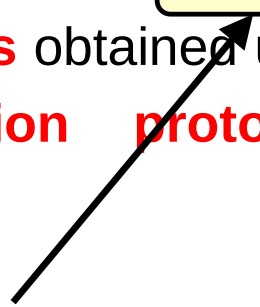


ISiS – Integrated Semantic Information and Knowledge System



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

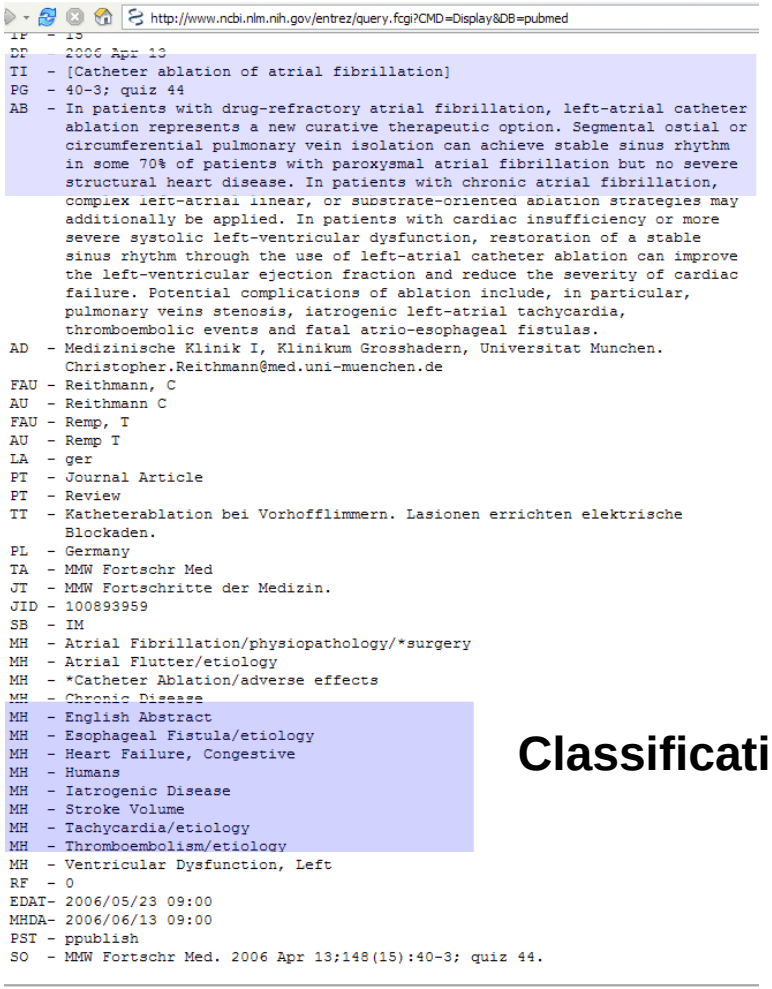
MeSH

- 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)

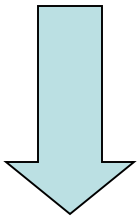


Abstract

Classification/Annotation

Method – Parse Sentences in PubMed

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

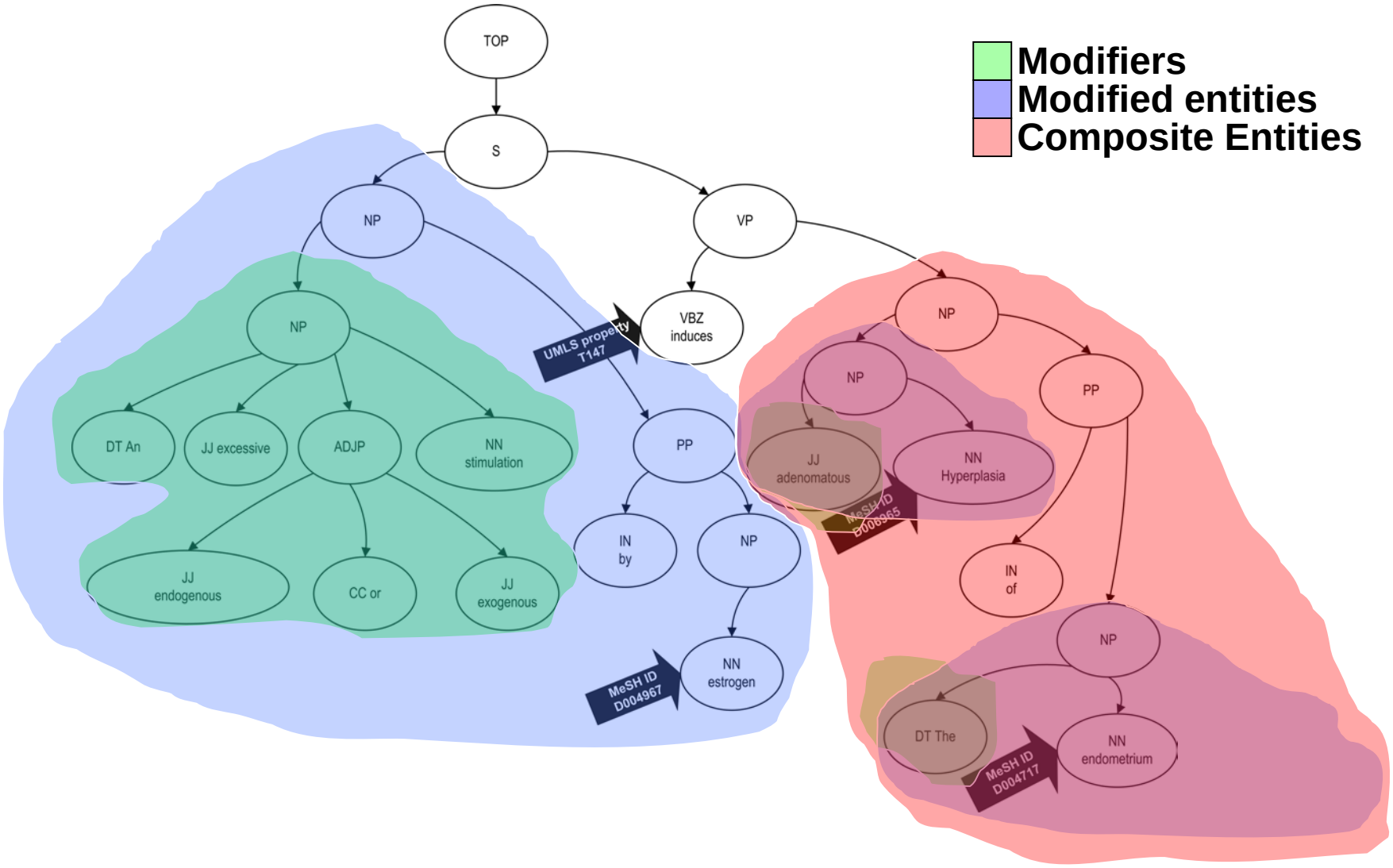


SS-Tagger (University of Tokyo)

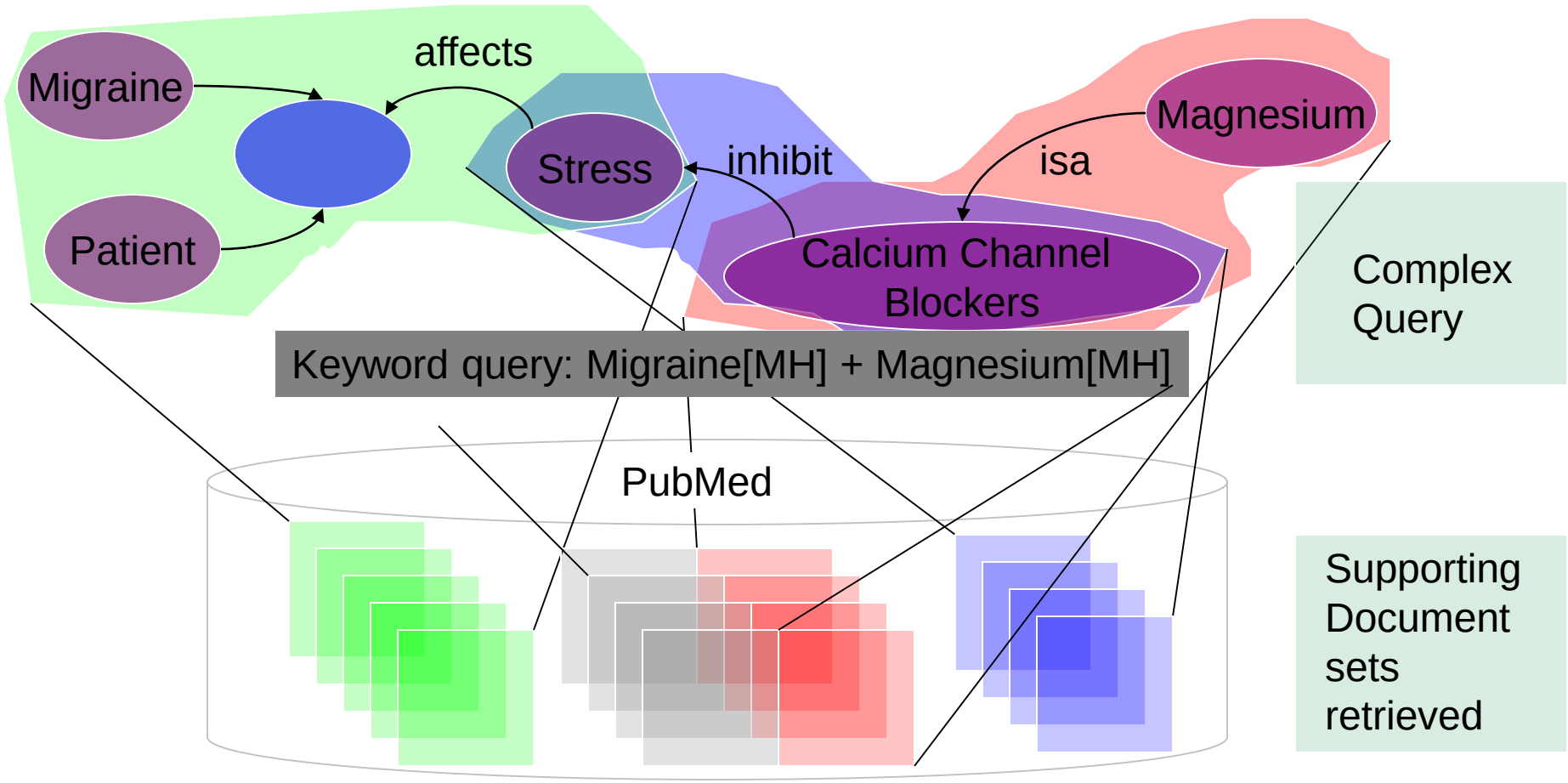
SS-Parser (University of Tokyo)

(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

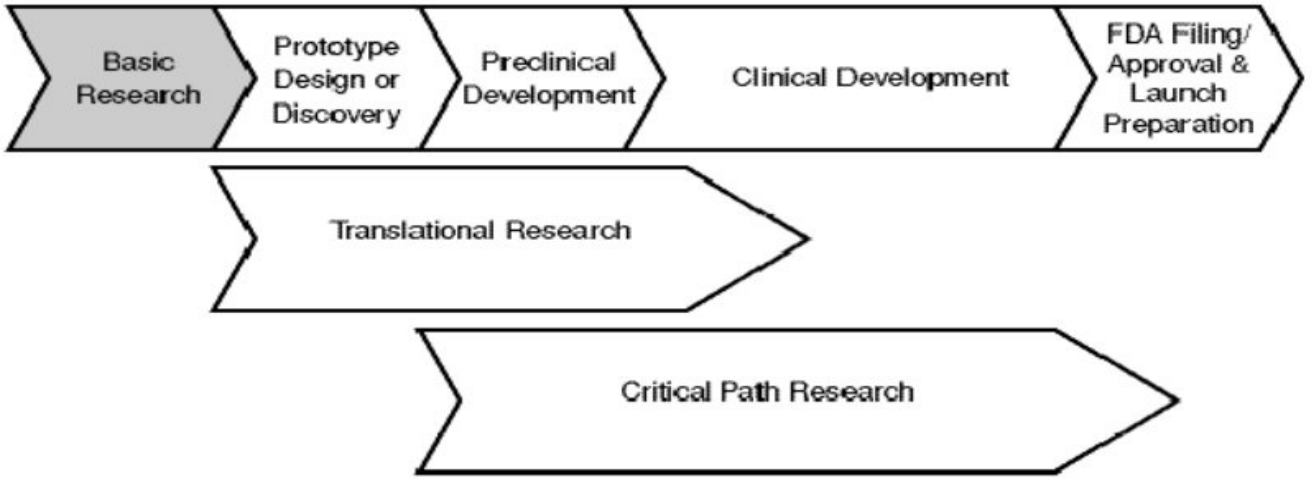
Advances in Science

- 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

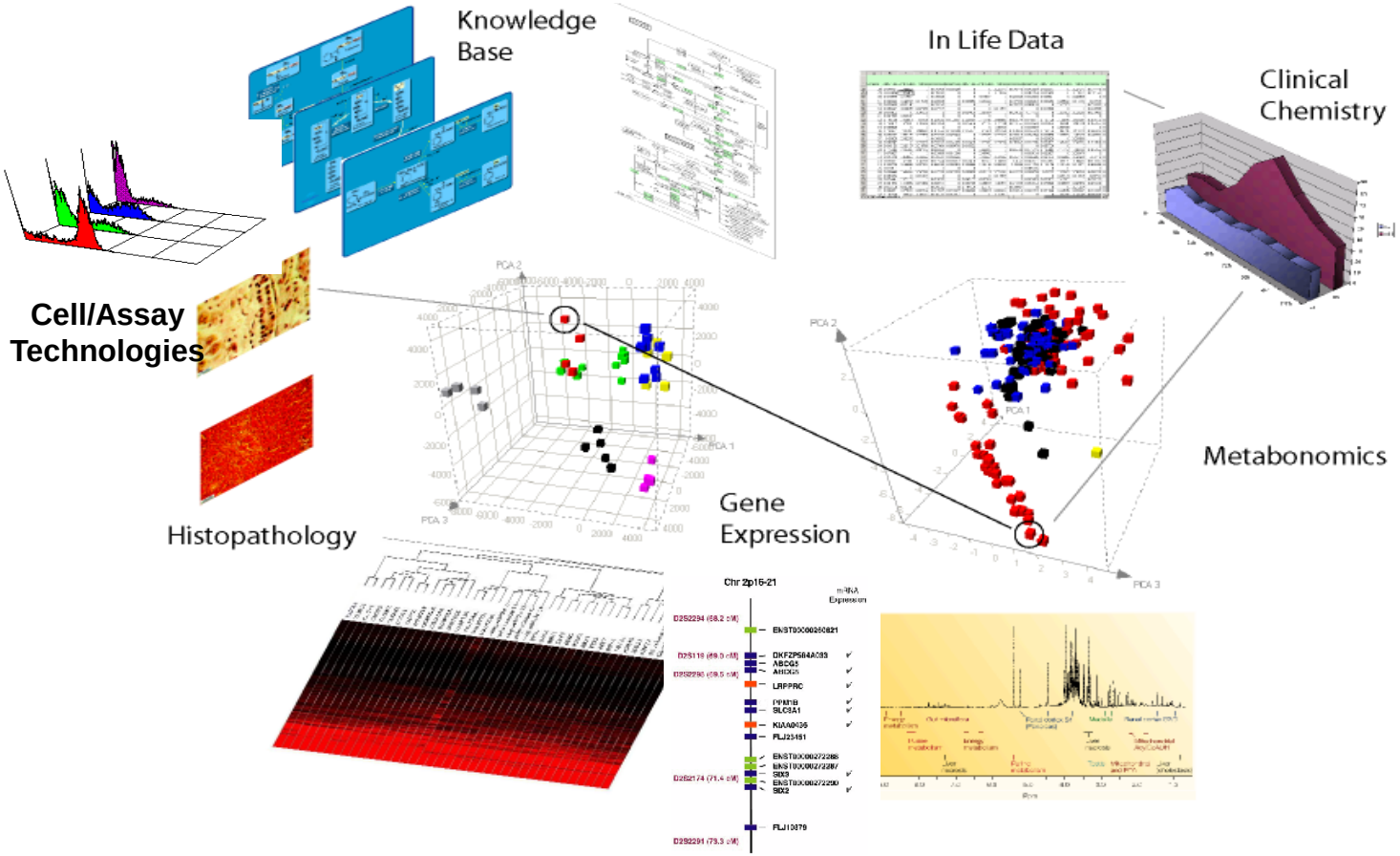
Critical Path Initiative



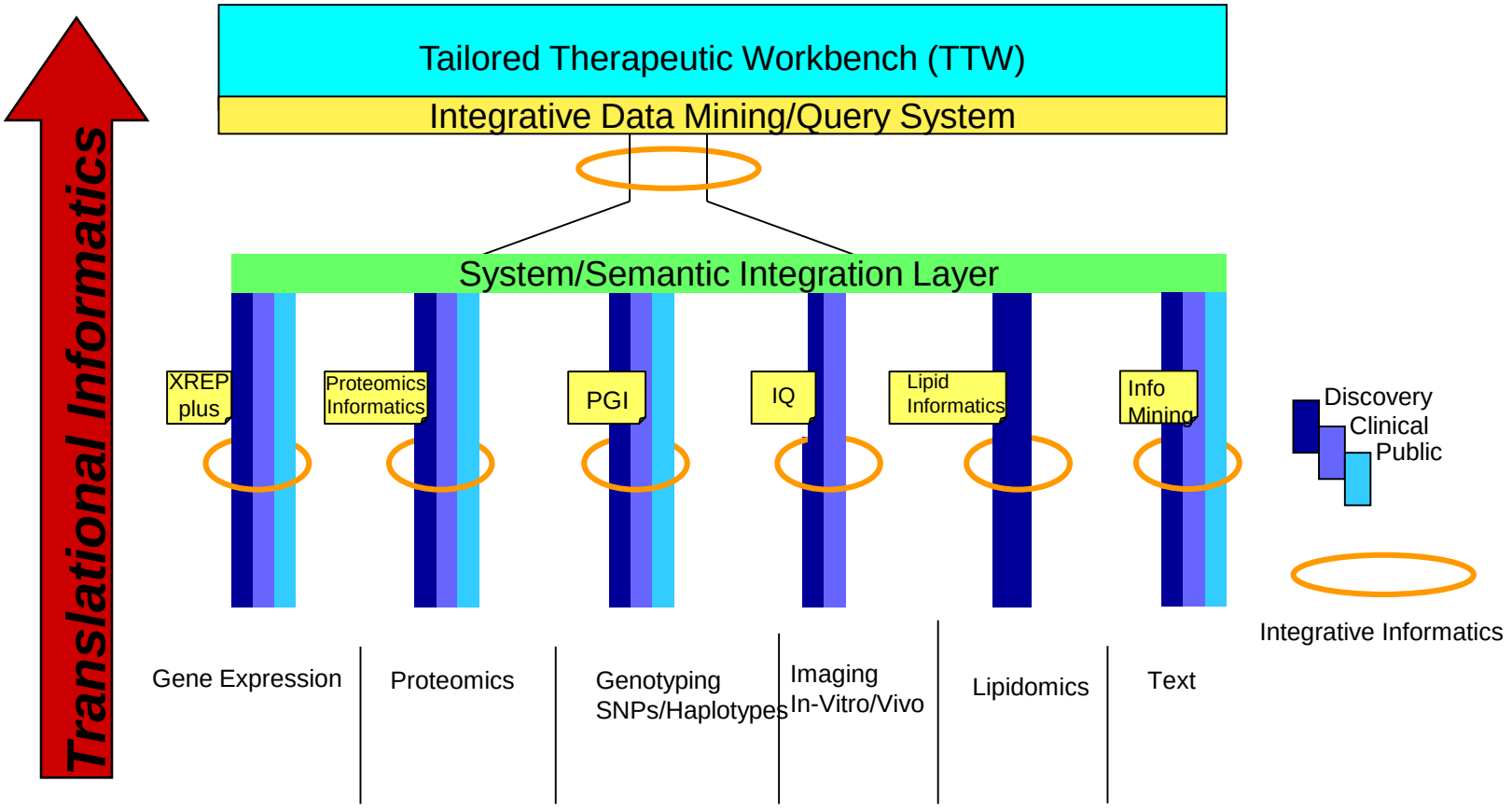
"In response to the widening gap between basic biomedical knowledge and clinical application, governments and the academic community have undertaken a range of initiatives. After decades of investment in basic biomedical research, the focus is widening to include *translational research* — multidisciplinary scientific efforts directed at "accelerating therapy development" (i.e., moving basic discoveries into the clinic more efficiently)."

Source: Innovation or Stagnation, FDA Report, March 2004

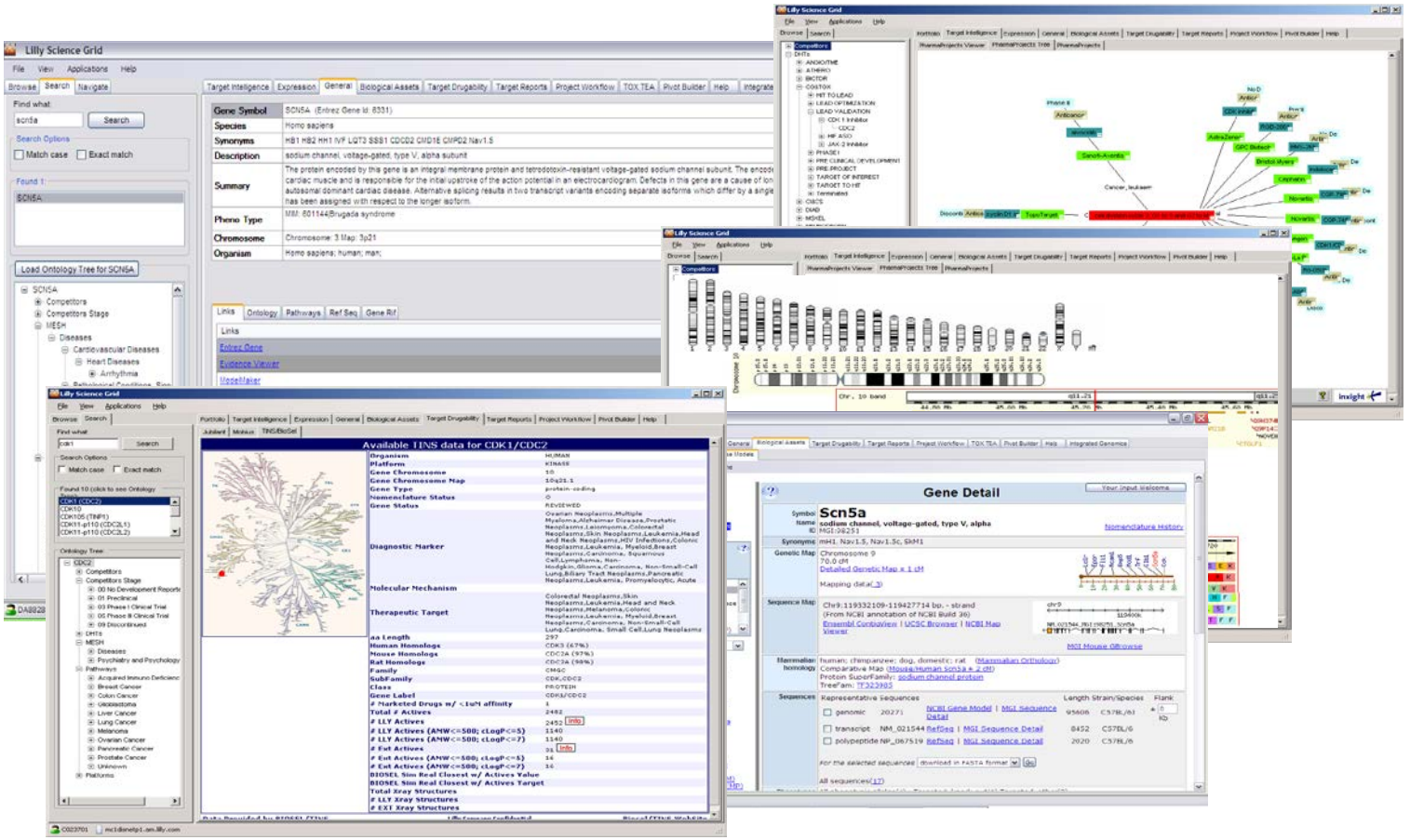
The Web of Heterogeneous Data



Integrative Informatics Platform



Discovery Target Assessment Tool



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

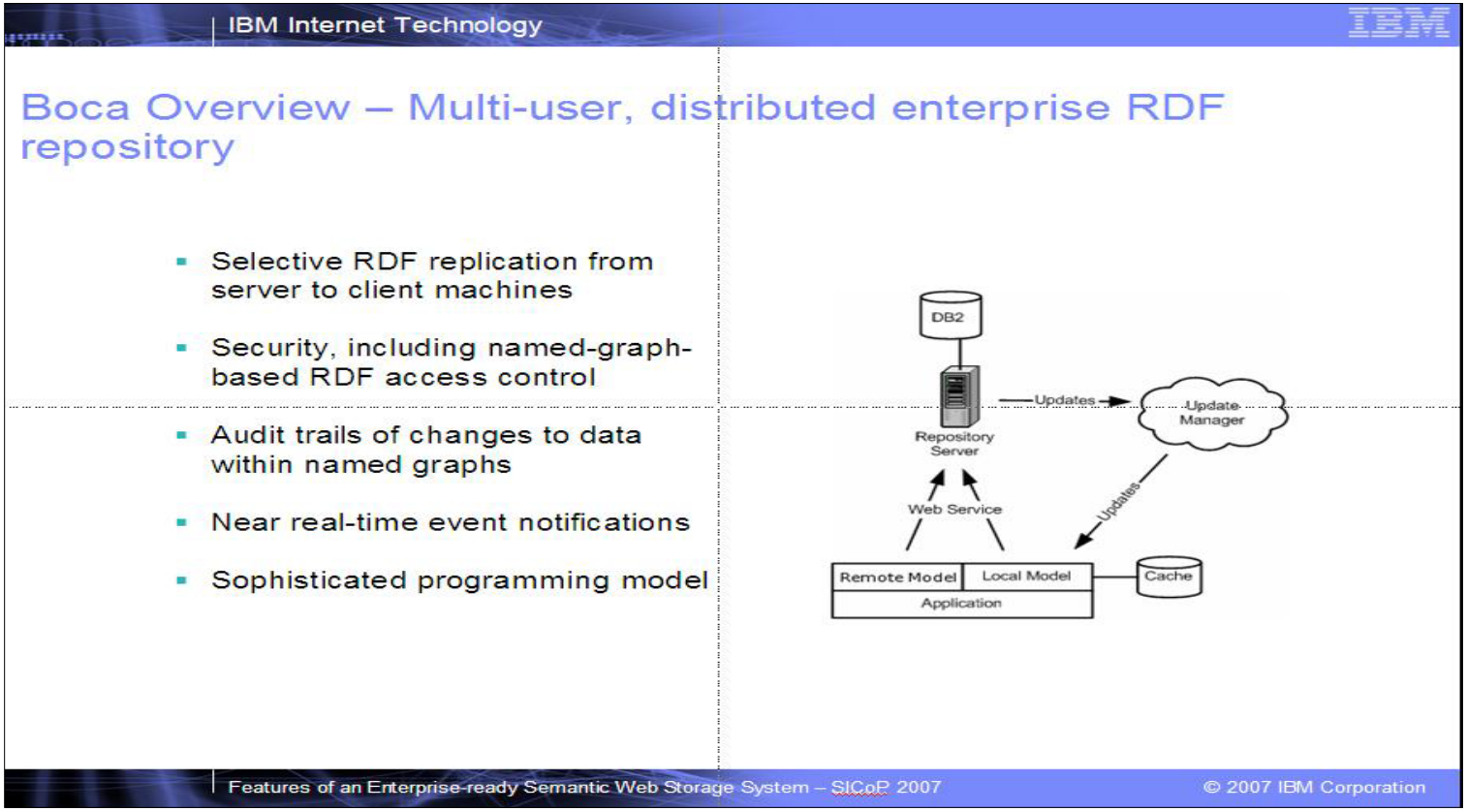
- Triple stores
- RDFizers
- Ontology Editors / Reasoning Systems
- Application Frameworks

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



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

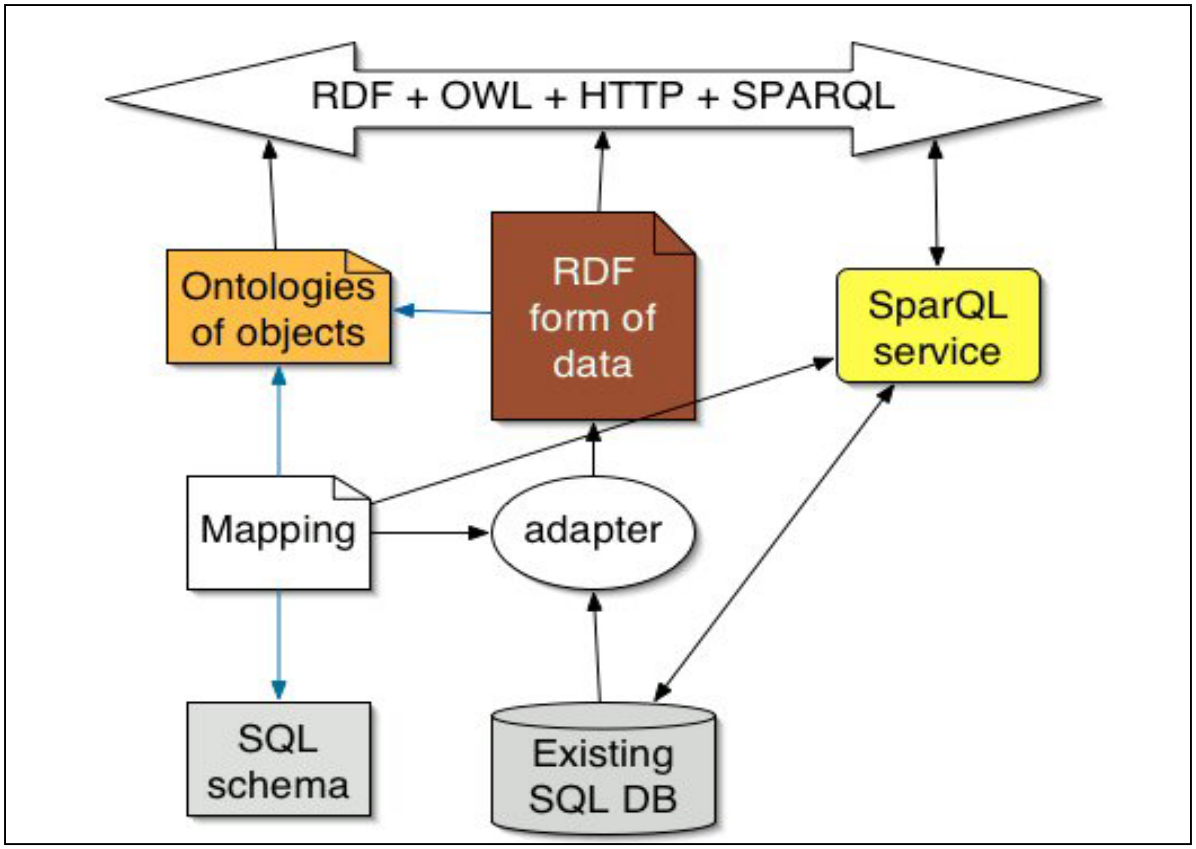
• Produces RDF Instance Data in Physical and Virtual forms from local or 3rd party data sources

• 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)

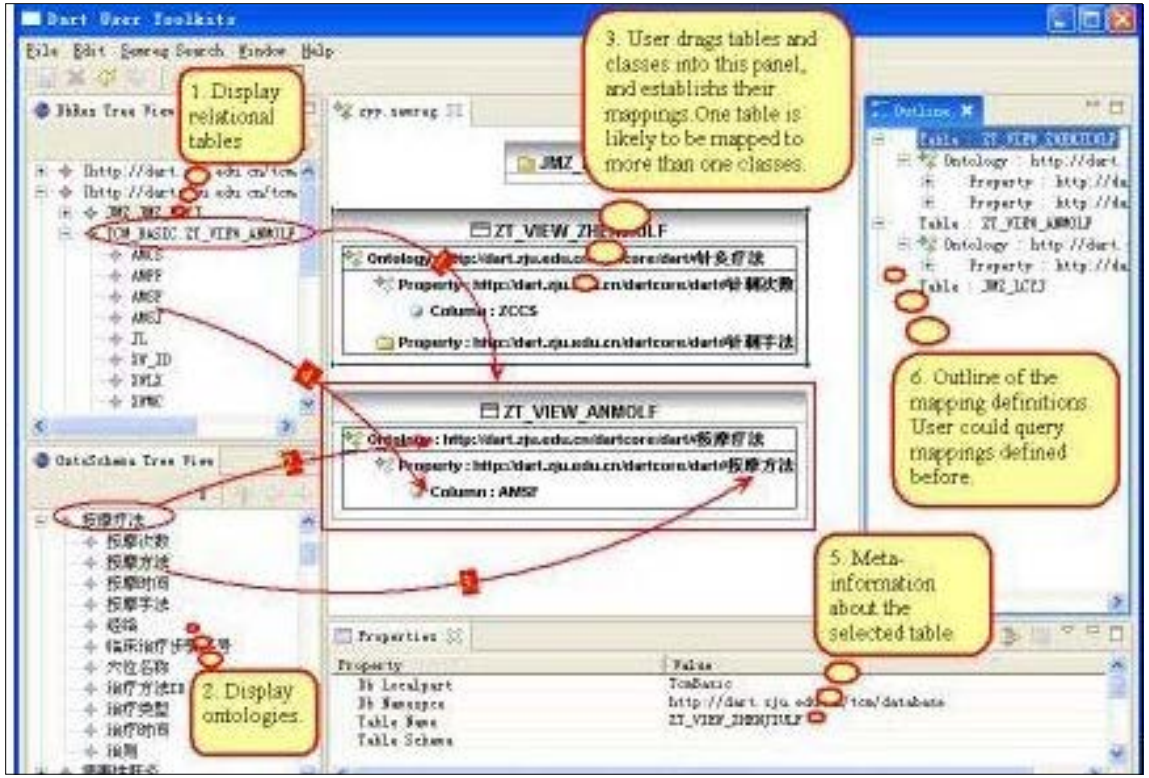
Adapting SQL Databases



Source: Tim Berners-Lee

Mapping Relational to RDF

- D2RQ
- SquirrelRDF
- DartGrid
- SPASQL



Source: DartGrid

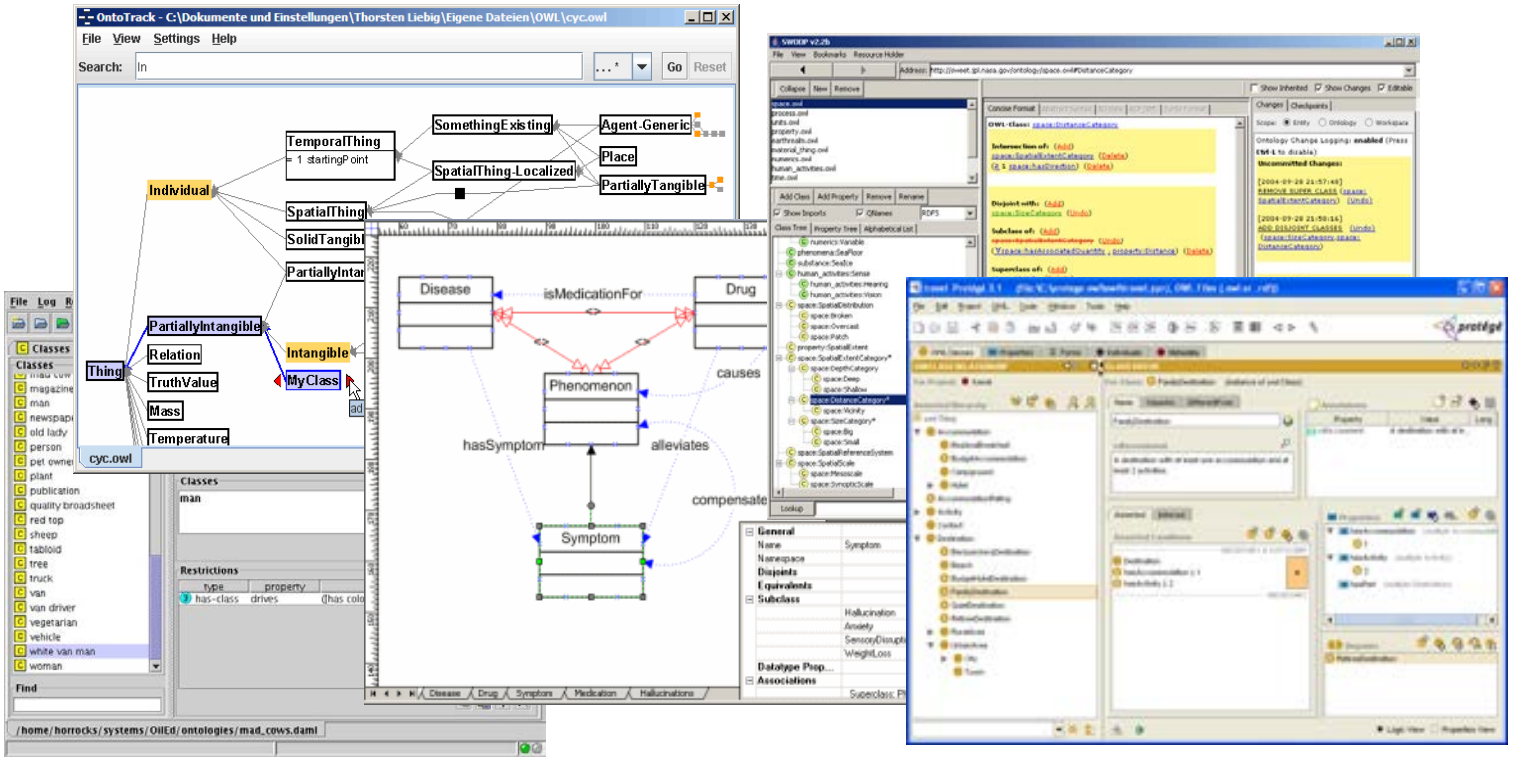
RDFizers

- 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

CEL

FaCT++

Racer

Pellet

KAON2

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
- 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