The Integration of Biological Data Using Semantic Web Technologies

Susie Stephens
Principal Product Manager, Life Sciences
Oracle

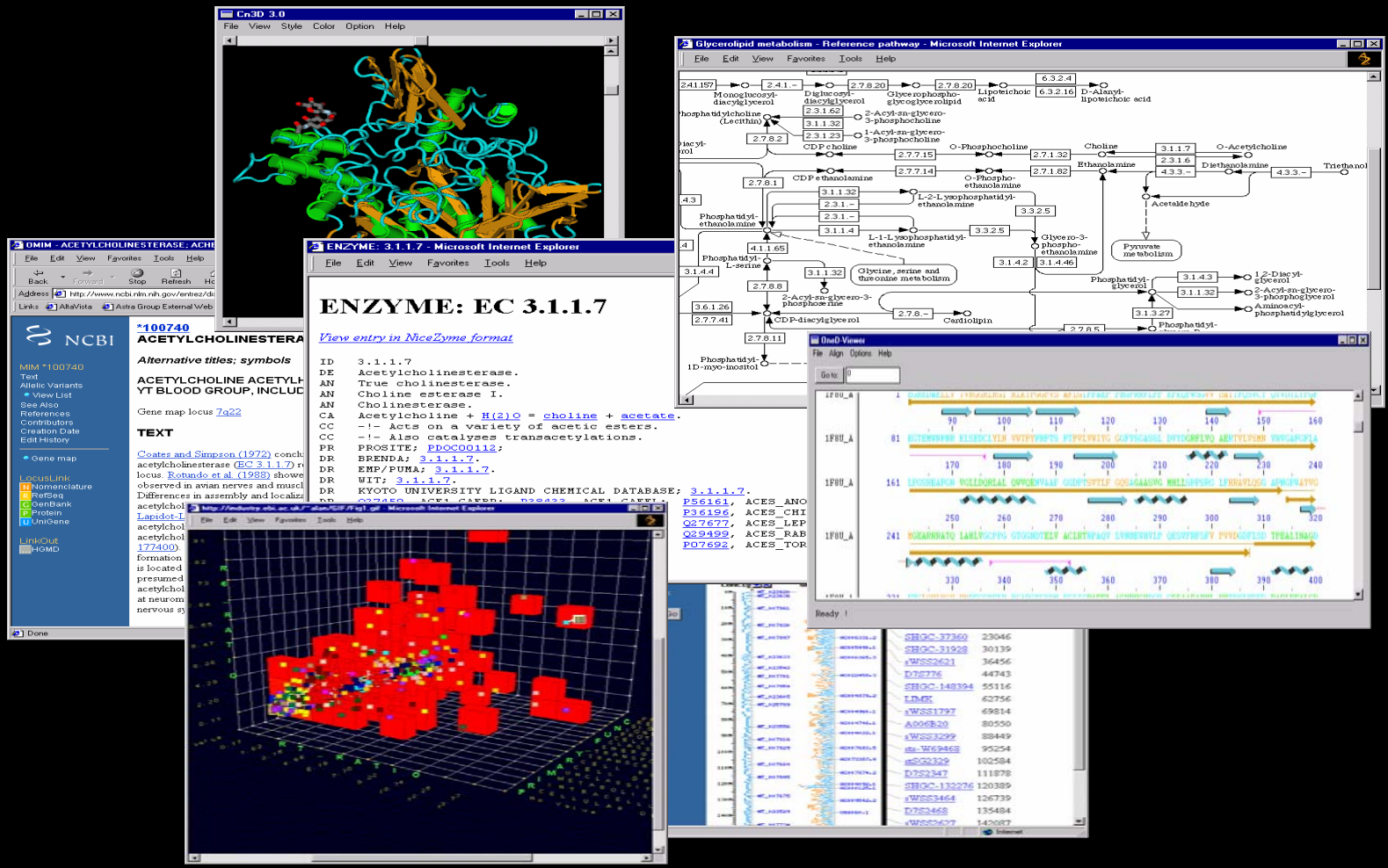
susie.stephens@oracle.com

ORACLE

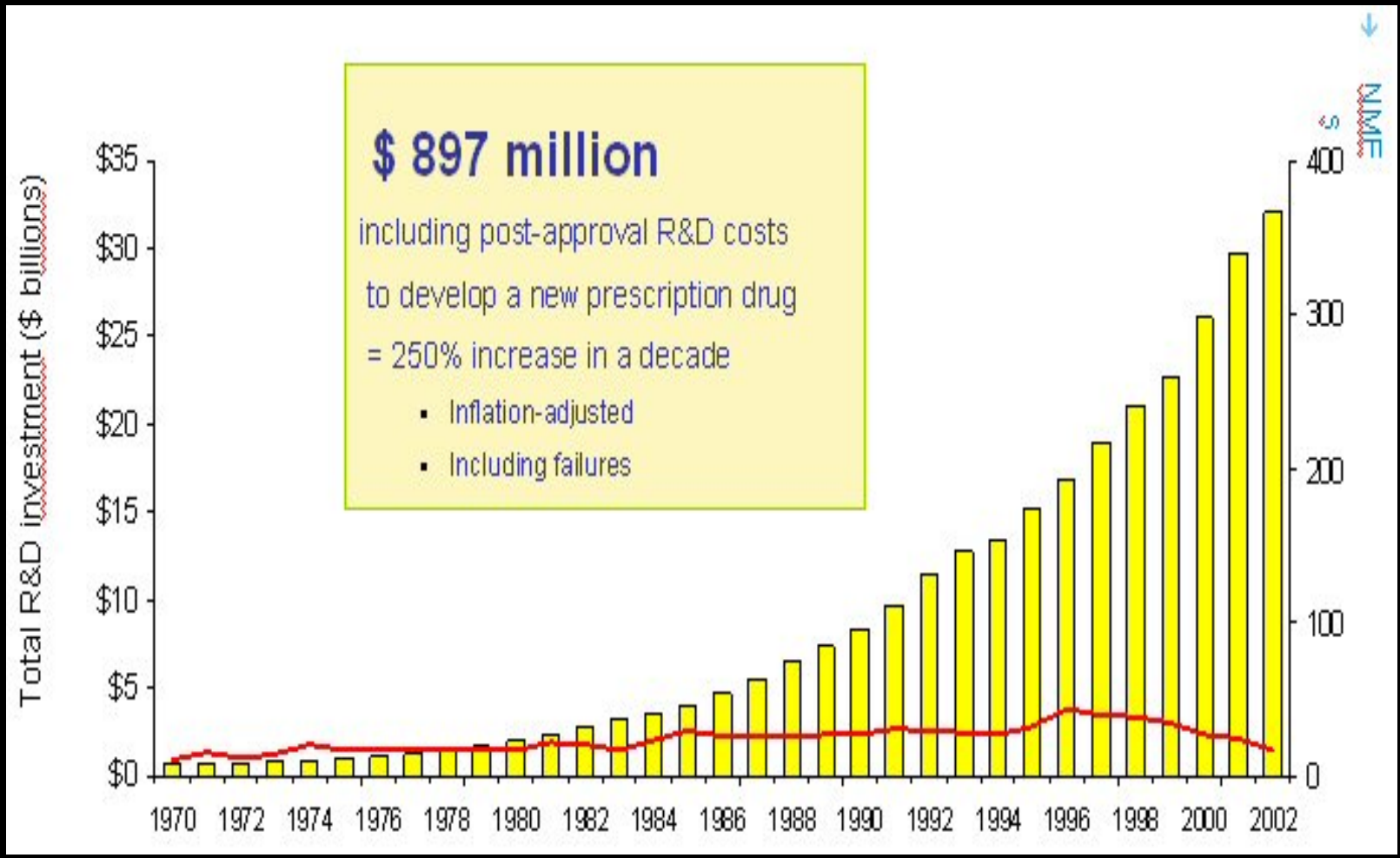
Outline

- Complexity of Biological Data
- Oracle's RDF Data Model
- Life Sciences Use Cases

The Complexity of Biological Data

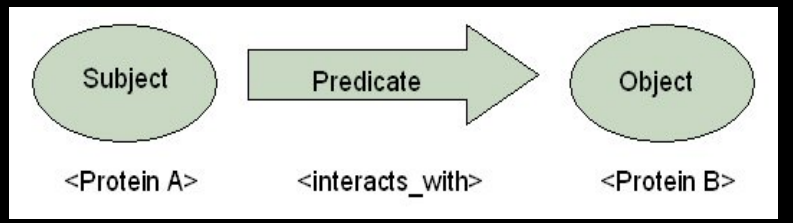


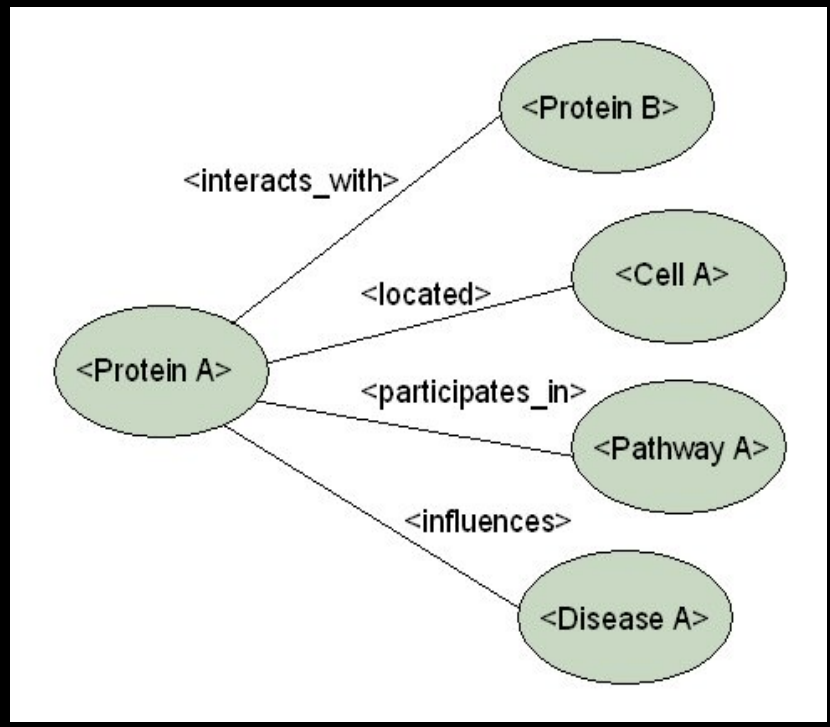
Pharmaceutical Productivity



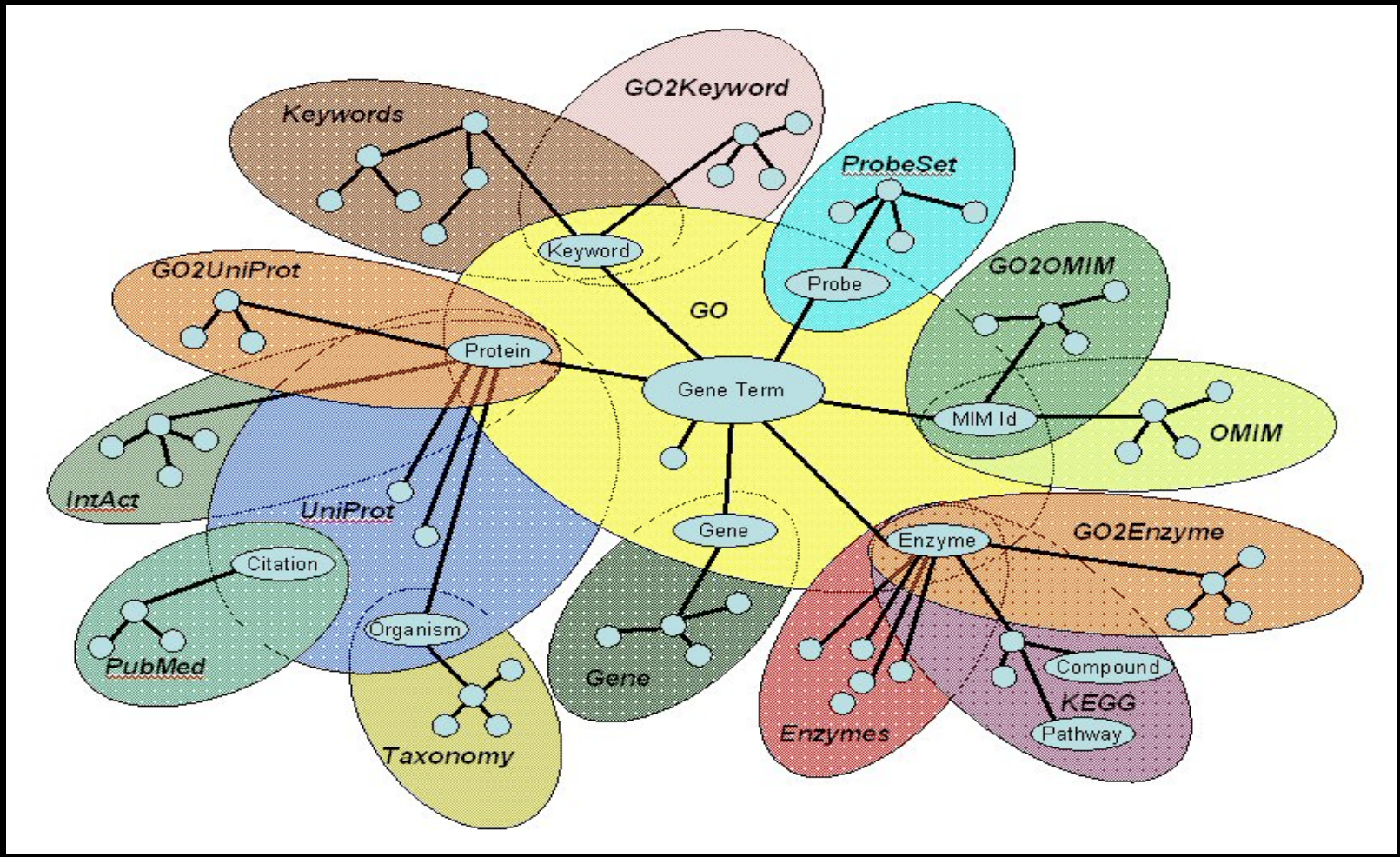
Source: PhRMA & FDA 2003

RDF Triples in Life Sciences





The Semantic Web Vision



Source: Stephens et al. J Web Semantics 2006

Outline

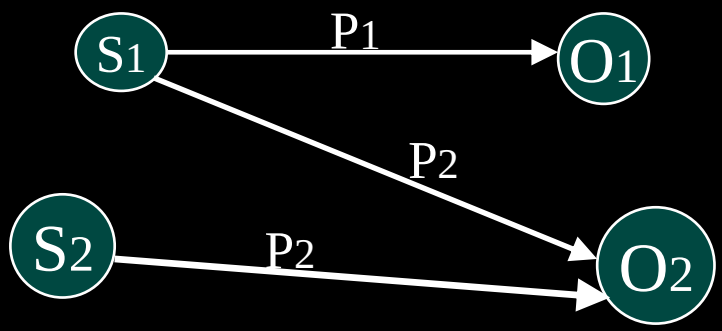
- Life Sciences Data
- Oracle’s RDF Data Model
- Use Cases

Oracle and RDF: Motivation

- Customer requests
- RDF (and OWL) are maturing
- Oracle supports open standards
- Complements Oracle's information management approaches
- Ability to leverage existing technologies

Oracle RDF Data Model

- Support for RDF and RDFS
- Object-relational implementation
- Subjects and objects are re-used
- Links represent complete RDF triples



RDF Triples:

- {S1, P1, O1}
- {S1, P2, O2}
- {S2, P2, O2}

SPARQL-like Query Capability

- A table function allows a graph query to be embedded in a SQL query
- Searches for an arbitrary pattern against the RDF data
- Includes inferencing based on RDF, RDFS, and user-defined rules

Enterprise Functionality

- Real Application Clusters (RAC), Security
- Multi-threaded, parallel processing, indexed, etc.
- Performance testing with UniProt

	Q1	Q2	Q3	Q4	Q5	Q6
10 M Triples	0.86	< 0.01	< 0.01	0.03	0.18	0.46
20 M Triples	0.95	< 0.01	< 0.01	0.03	0.19	0.47
40 M Triples	0.96	< 0.01	< 0.01	0.03	0.18	0.47
80 M Triples	1.03	< 0.01	< 0.01	0.03	0.20	0.49
Maximum σ	.054	0.002	0.002	.011	.065	0.07

Units in seconds

Source: Chong et al. VLDB 2005

Image Search

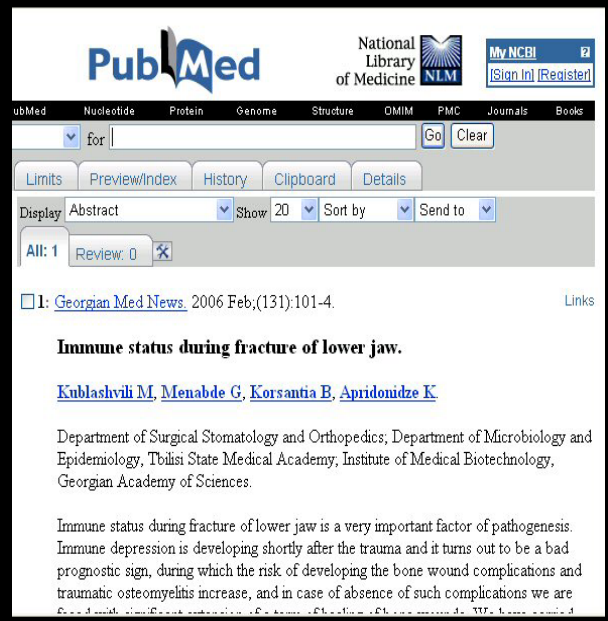
“Find me all DICOM images that contain the term ‘Jaw’”

- Map relationships to terms using RDF triples
- ‘Mandible’, sameAs’, ‘Jaw’
- ‘Maxilla’, ‘partOf’, ‘Jaw’

Text Search

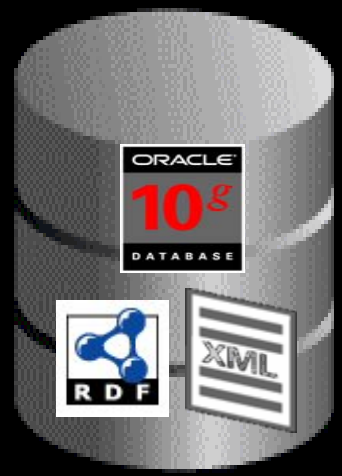
“Find me all papers that contain the term ‘Jaw’”

- Map relationships to terms using RDF triples
- ‘Mandible’, sameAs’, ‘Jaw’
- ‘Maxilla’, ‘partOf’, ‘Jaw’



Data Integration

- SQL / RDBMS
  - Concise, efficient transactions
  - Transaction metadata is embedded or implicit in the application or database schema
- XQuery / XML
  - Transaction across organizational boundaries
  - XML wraps the metadata about the transactior around the data
- SPARQL / RDF
  - Information sharing with ultimate flexibility
  - Enables semantics as well as syntax to be embedded in documents



Download the Database!

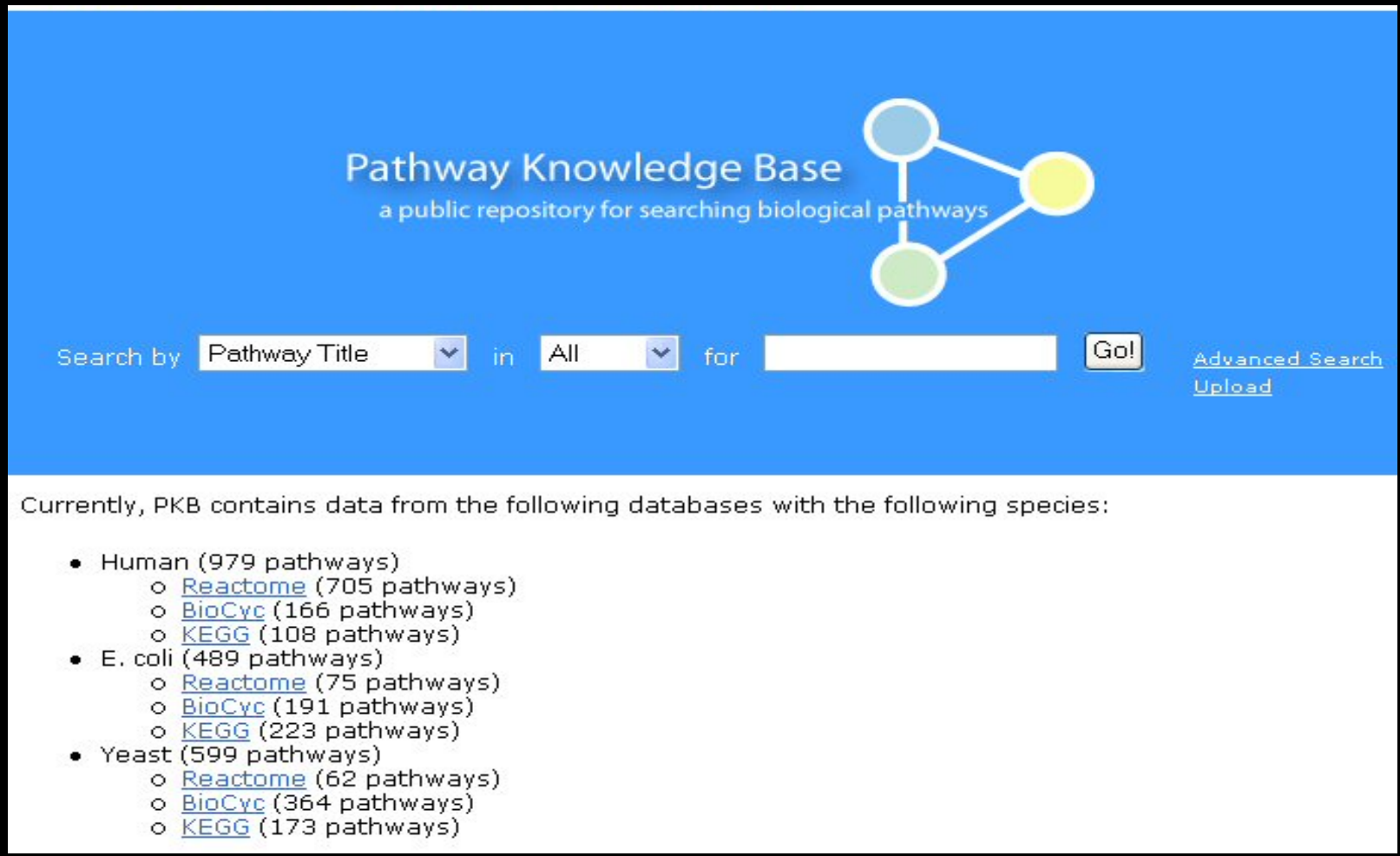
Oracle Database Enterprise Edition 10g Release 2

http://www.oracle.com/technology/software/products/database/oracle10g/index.html

Outline

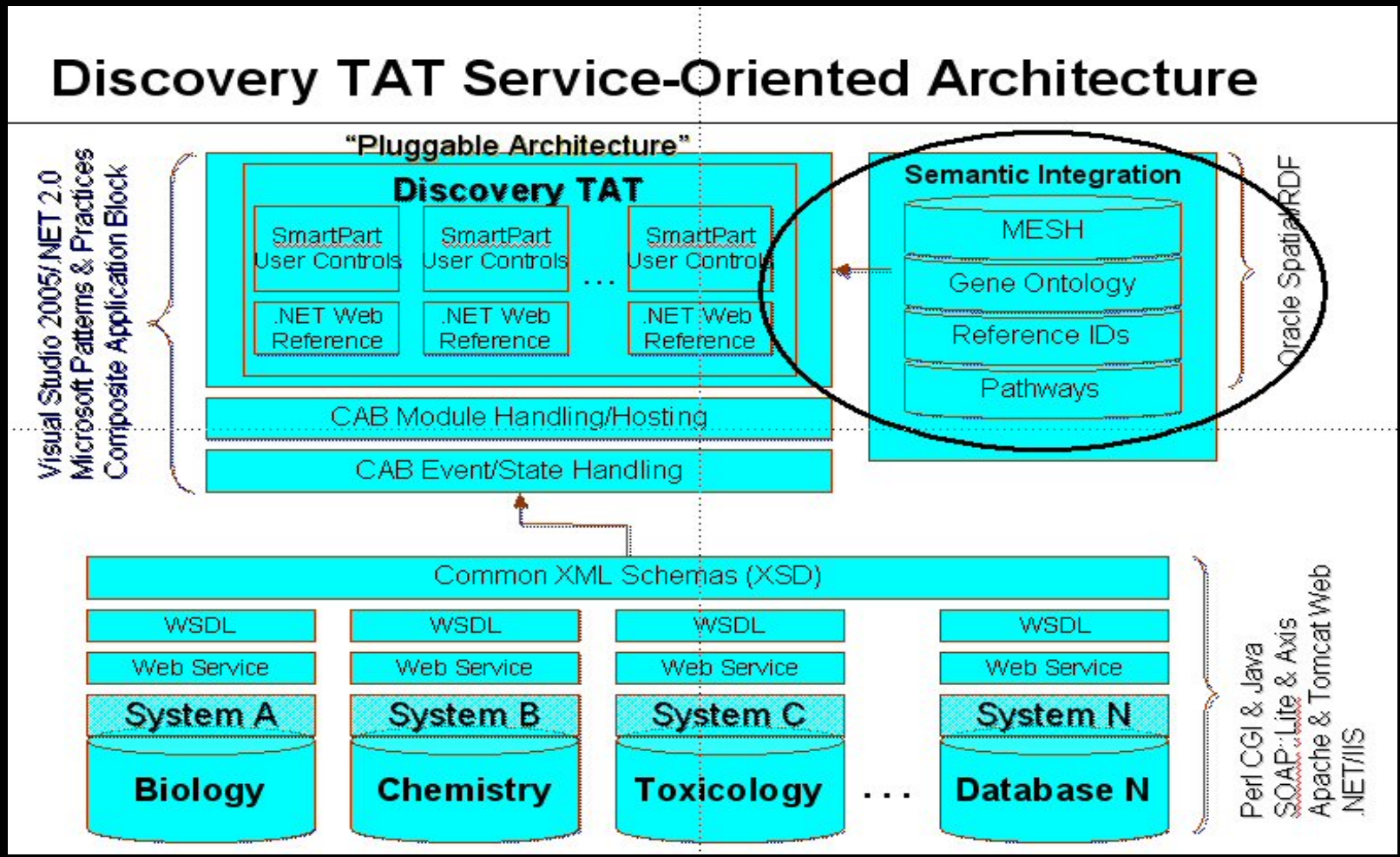
- Life Sciences Data
- Oracle’s RDF Data Model
- Use Cases

Stanford University Use Case



Source: http://pkb.stanford.edu/

Eli Lilly Use Case



Source: http://www.olsug.org/wiki/images/d/df/AWL.pdf

University of Texas Health Science Center Use Case

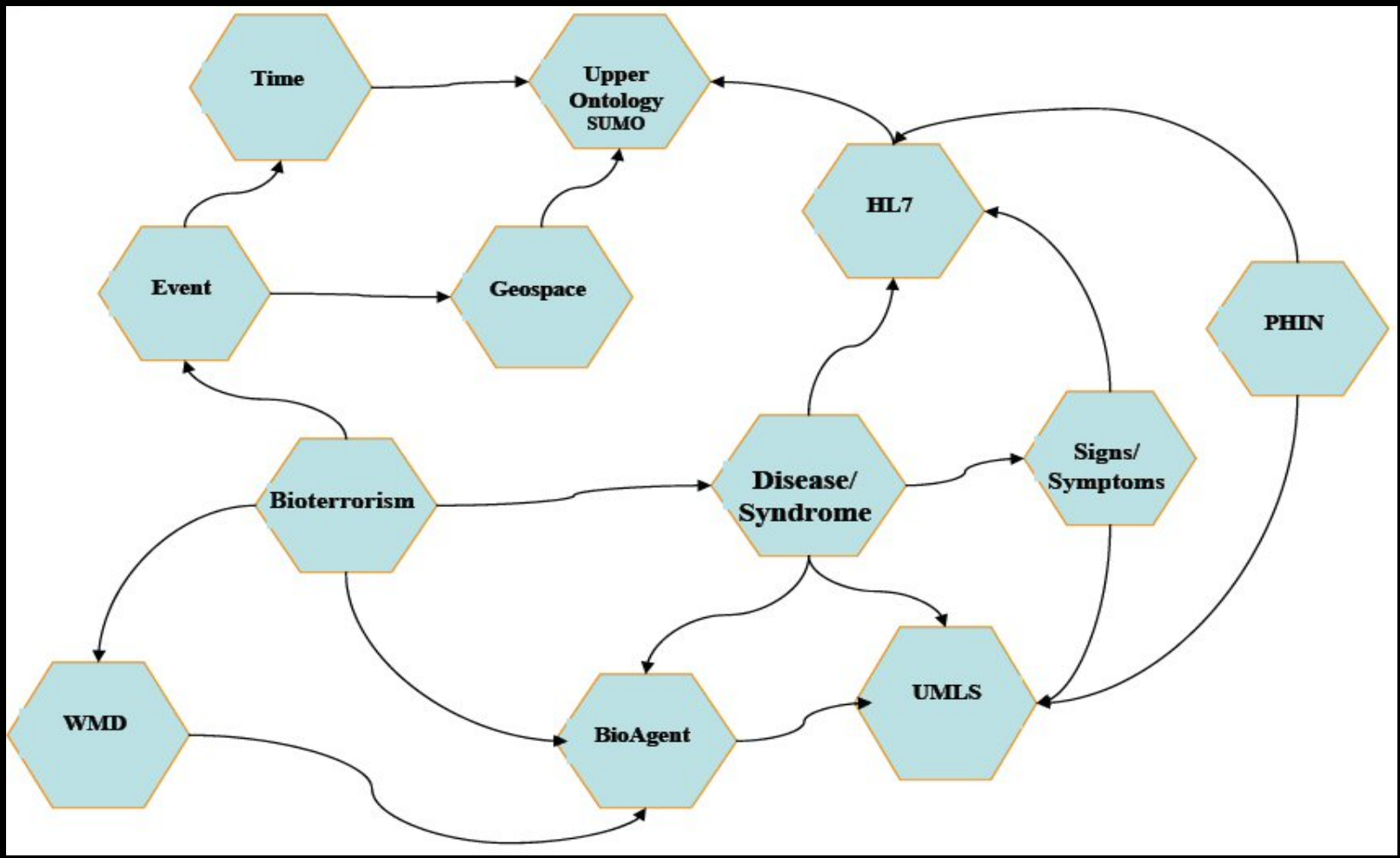
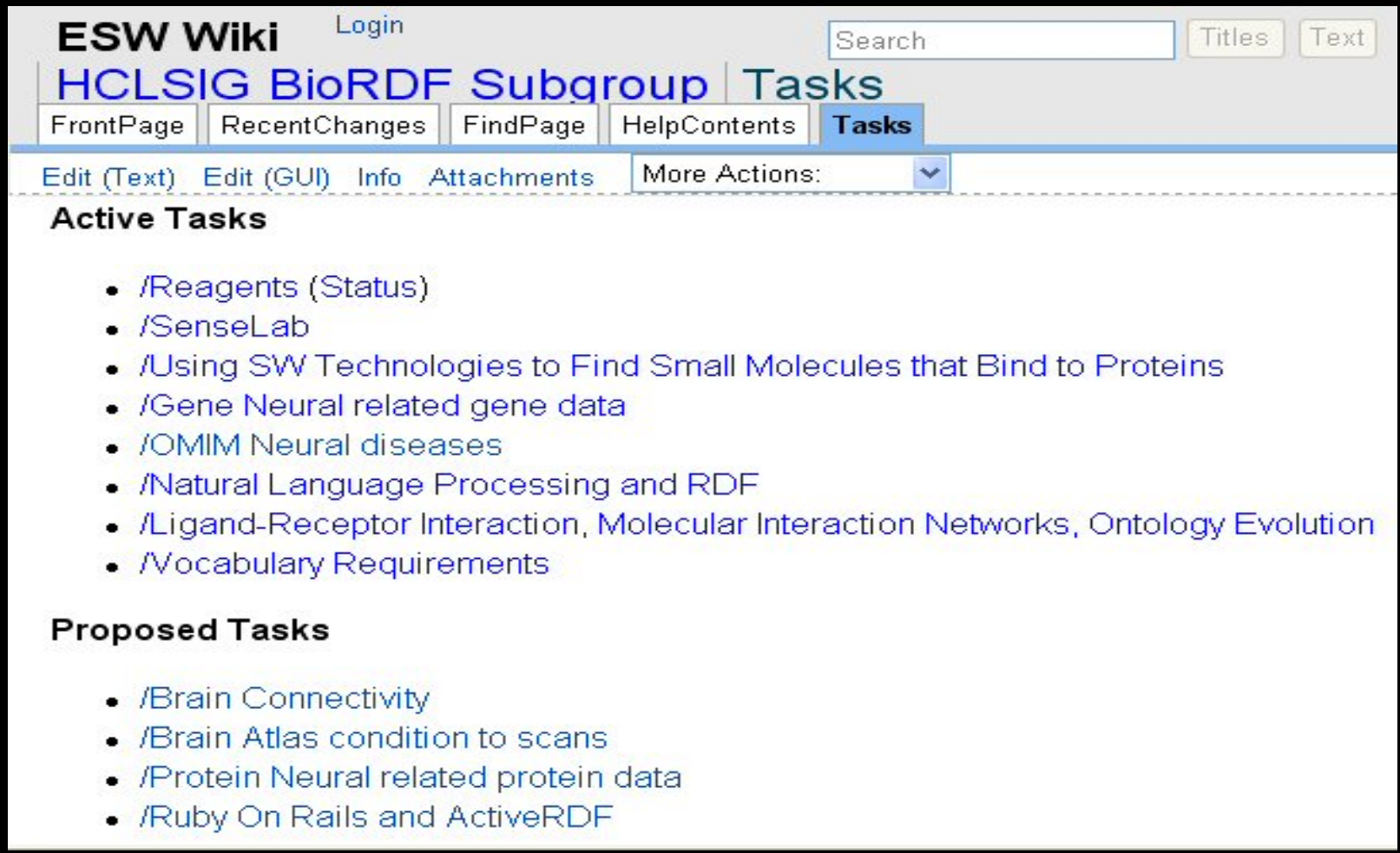


Image Source: Semantic Technologies Conference 2006

BioRDF



Source: http://esw.w3.org/topic/HCLSIG_BioRDF_Subgroup

Summary

- The Semantic Web provides the ability to more easily integrate heterogeneous data
- Oracle has a scalable, secure, highly-available RDF Data Model
- Adoption of Semantic Web technologies is accelerating
- Make your data sharable, make it available in RDF