CHEMICAL SEMANTICS

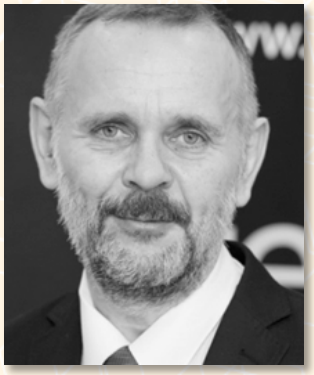
Semantic Web for Chemistry

- Chemical Semantics, Inc. – A Startup In Florida
- Neil Ostlund, CEO
- Mirek Sopek, President
- Application to Computational Chemistry

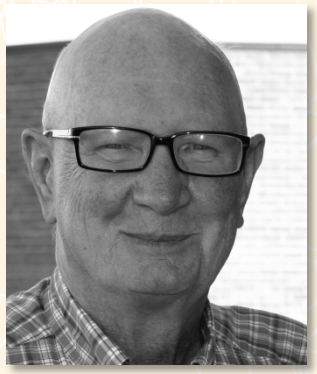
CHEMICAL SEMANTICS

Who are the people?

Two Ph.D. theoretical chemists who are close friends and have extensive experience in computer science and running successful software companies.



Mirek Sopek, founder, 1989, of MakoLab S.A. (www.makolab.pl). MakoLab is a public company in Poland with extensive experience in web design, business software, and recently the semantic web.



Neil Ostlund, founder, 1985, of Hypercube, Inc. (www.hyper.com). Hypercube, is a Canadian Company, that is a leader in molecular modeling software. It has dealers in 35 countries selling its HyperChem product.

50 years running software companies.

CHEMICAL SEMANTICS

Chemical Semantics

Applying the Semantic Web To Computational Chemistry

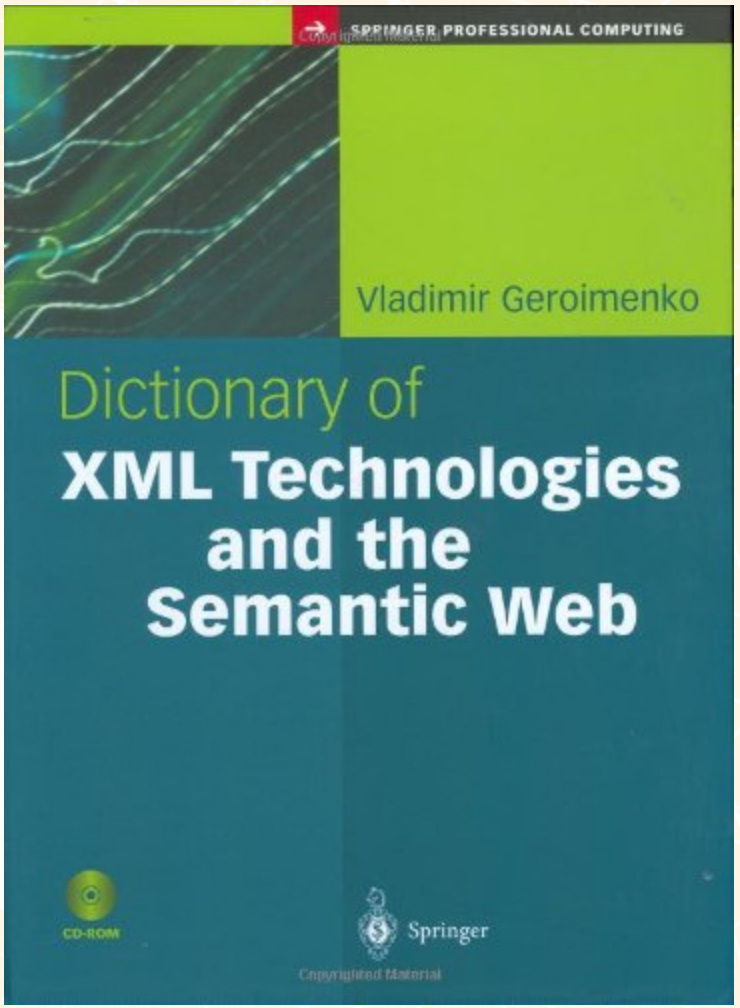
What are the Principal Ideas?

CHEMICAL SEMANTICS

What is Semantic Web?

- Web 3.0
- Web of Data
- Web of Meaning
- RDF, OWL, SPARQL

An extension of the current Web that provides an easier way to find, share, reuse and combine information. It is based on machine-readable information and builds on XML technology's capability to define customized tagging schemes and RDF's (*Resource Description Framework*) flexible approach to representing data. The Semantic Web provides common formats for the interchange of data (where on the Web there is only an interchange of documents)



CHEMICAL SEMANTICS

WWW Generations

- WWW 1.0 – **The Original** - read only
- WWW 2.0 – **The Social Web**
  Example Facebook, YouTube
  - read/write
- WWW 3.0 – **The Semantic Web** – read/write/execute

Semantic Web is a group of methods and technologies to allow machines to understand the meaning - or "semantics" - of information on the World Wide Web.

Technologies – Resource Description Framework (RDF) - Data
- Web Ontology Language (OWL) – Meaning of Data
- SPARQL Protocol and RDF Query Language

CHEMICAL SEMANTICS

What is CSI principal idea?

- Create **Publishing Software** that allows scientists, principally computational chemists and biochemists, to publish their **data** in a semantic (knowledge base) fashion.

- Create **Agent Software** that allows these scientists to search the world for relevant scientific results using **inference to create new results** from the existing global data.

Is the idea any good?

The established system of journals for communicating the results of scientific research is already being challenged by the existence of the web. But we are only in the early days of a **new Internet revolution**, one which will have a deeper and more disruptive impact on scientific, and other, web publishing, and have profound implications for the web itself.

An emerging successor to the web, **the Semantic Web**, will likely profoundly change the very nature of how scientific knowledge is produced and shared, in ways that we can now barely imagine.

T Berners-Lee, J. Hendler – Nature, 2001

CHEMICAL SEMANTICS

Is the idea any good?

Sir Tim Berners-Lee Gives Congress Vision Of The Future, Information Week, 2007

"Science thrives when ideas, hypotheses, data, and knowledge are quickly and easily shared within disciplines and communities. The advent of the Internet and Web accelerated information sharing with tools such as e-mail, online publishing, digital libraries, and comprehensive search engines such as Google.

Researchers and developers are now exploring a new idea that many believe will further enhance scientists' ability to share knowledge: the **Semantic Web**."

Congressional subcommittee hearing, entitled The Digital Future of the United States, *T Berners-Lee.*

CHEMICAL SEMANTICS

Is the idea any good?

semanticweb.com™ The Voice of Sema and Linked Data

Home | Events | Community | Learning | Industry Verticals | Answers | Jobs

BIG DATA, ENTERPRISE DATA

Gartner Names Semantic Technologies To Its Top Technology Trends Impacting Information Infrastructure in 2013

By Jennifer Zaino on March 7, 2013 10:53 AM

LISTED

Semantic technologies have made it to Gartner's list of the top technology trends that will impact information infrastructure this year.

The research firm yesterday released the list of nine trends that it says will play key roles in modernizing information management and in making the role of information governance increasingly important. Semantic technologies come in at No.3 on the list – right behind closely-tied-to trends Big Data and modern information infrastructure.

CHEMICAL SEMANTICS

Scientific Publishers

- Inaccessible Data
- Take Too Long
- Have Archaic Tools
- Charge Too Much

"A revolution, then, has begun. Technology permits it; researchers and politicians want it. If scientific publishers are not trembling in their boots, they should be." The Economist.

CHEMICAL SEMANTICS

Fundamental Technologies

XML – Extensible Markup Language

XML – a good way to structure data

XML – tree structured

XML – precursor to Semantic Web?

XML – not semantic web however

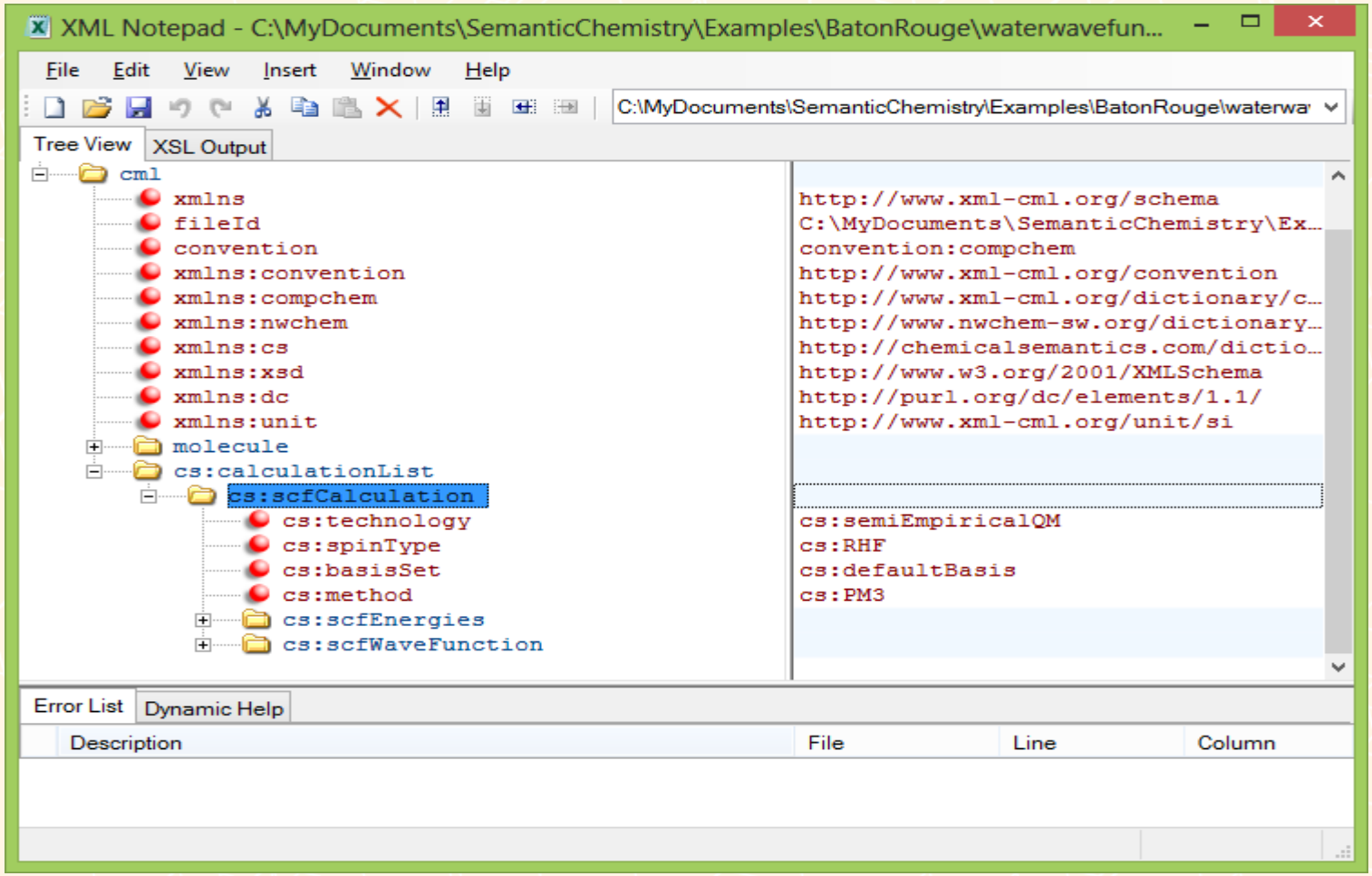
Markup used for more than Semantic Web

CML – Chemical Markup Language
- Murray-Rust and the English Mafia (:>)

CSX – Chemical Semantics XML - test bed

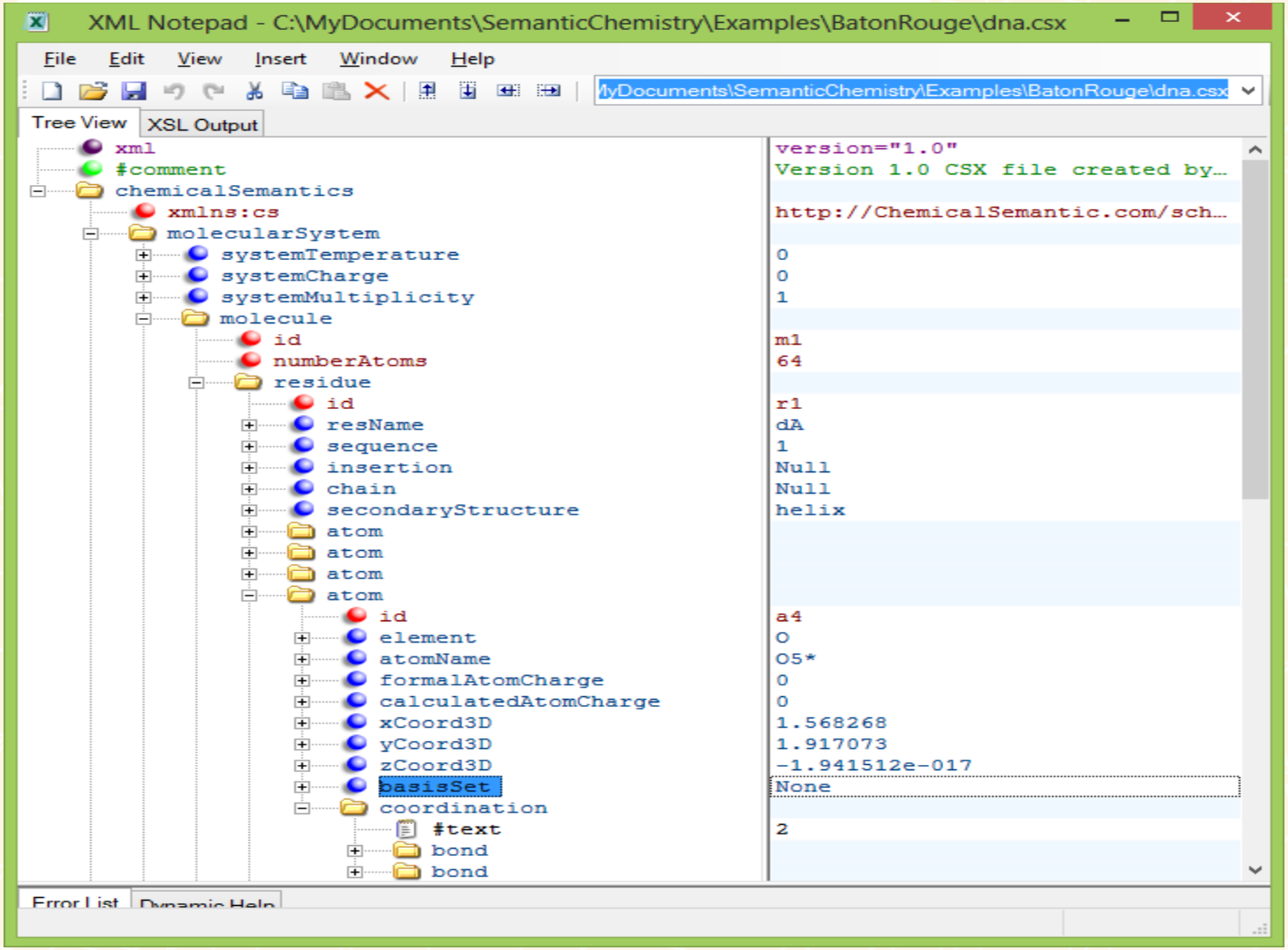
CHEMICAL SEMANTICS

CML (CS dictionary)



CHEMICAL SEMANTICS

CSX (proteins and dna)



CHEMICAL SEMANTICS

Fundamental Technologies

Uniform Resource Indicator (URI) – name data

Resource Description Framework (RDF) – hold data

Web Ontology Language (OWL) – describe data

SPARQL Protocol and RDF Query Language – find data

Each of these is a standard set by World Wide Web Consortium (W3C) and is a component of the semantic web.

Linked Data a subset of Semantic Web with less emphasis on Ontology.

Been around for a decade but still essentially unimplemented.

IT'S TIME HAS COME!

URI

Generalization of URL

An unambiguous name for a “resource” (thing, idea, etc.)

e.g. http://cs.com/neil/dogs/lucy

(not a page about Lucy but Lucy!)

e.g. http://cs.com/chem/dictionary/scfTotalEnergy

e.g. http://cs.com/chem/molecules/ethane

e.g. http://cs.com/chem/structure/hasBond

RDF

Graphical Database

(subject, predicate, object)

Subject is a unique resource named by URI

Predicate is unique resource named by URI

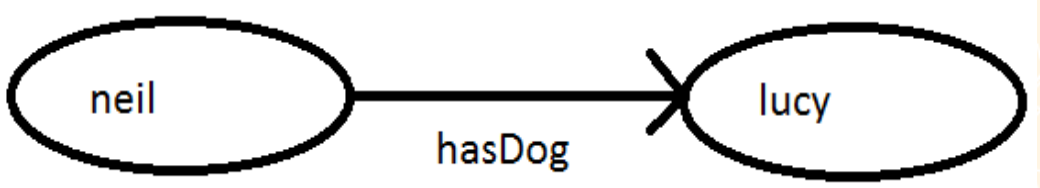
Object is literal or a resource named by URI

(neil, hasDog, lucy)

Subject = <http://cs.com/neil>

Predicate = <http://cs.com/vocabulary/hasDog>

Object = <http://cs.com/neil/dogs/lucy>



RDF

TRIPLES

http://cs.com/chem/molecules/water

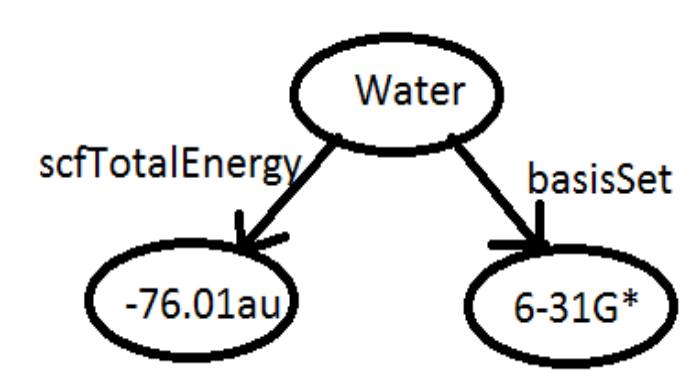
http://cs.com/chem/dictionary/scfTotalEnergy

“-76.01”

http://cs.com/chem/molecules/water

http://cs.com/chem/dictionary/basisSet

http://cs.com/chem/dictionary/6-31g*



CHEMICAL SEMANTICS

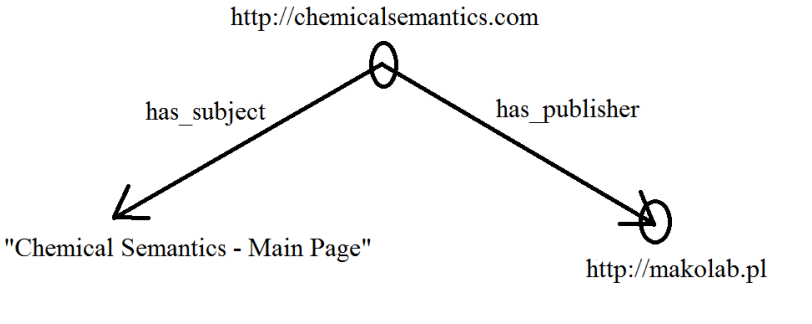
RDF

RDF Triples

Subject: http://chemicalsemantics.com (URI, Uniform resource indicator)

Predicate: has_publisher (defined by Dublin Core)

Object: http://makolab.pl (literal or URI)




```
<?xml version="1.0"?>
< rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc= "http://purl.org/dc/elements/1.1/">
< rdf:Description rdf:about="http://chemicalsemantics.com">
<dc:subject>Chemical Semantics - Main Page</dc:subject>
<dc:publisher>"http://makolab.pl</dc:publisher>
< /rdf:Description>
< /rdf:RDF>
```

Wikipedia

By encouraging the inclusion of semantic content in web pages, the Semantic Web aims at converting the current web dominated by unstructured and semi-structured documents into a "web of data".

CHEMICAL SEMANTICS

RDF Serialization – preliminary example

RDF/XML or

Turtle (Terse RDF Triple Language)

```
1  @prefix cs: <http://ChemicalSemantics.com/chem/dictionary/ns#> .
2  @prefix mol: <http://ChemicalSemantics.com/chem/molecules/simplewater.ttl#> .
3  @prefix xs: <http://www.w3.org/2001/XMLSchema#> .
4  mol:molecule_31 a cs:molecule ;
5  cs:name “water" ;
6  cs:atom _:atom31_1 ;
7  cs:atom _:atom31_2 ;
8  cs:atom _:atom31_3 ;
9  cs:bond _:bond31_1 ;
10 cs:bond _:bond31_2 .
11  _:atom31_1 cs:atomType cs:O ;
12  cs:x3 "-0.381950"^^xs:double;
13  cs:y3 "0.243825"^^xs:double;
14  cs:z3 "0.000000"^^xs:double.
```

CHEMICAL SEMANTICS

RDF Serialization

```
15  _:atom31_2 cs:atomType cs:H ;
16  cs:x3 "-0.381950"^^xs:double;
17  cs:y3 "1.203825"^^xs:double;
18  cs:z3 "0.000000"^^xs:double.
19  _:atom31_3 cs:atomType cs:H ;
20  cs:x3 "0.523148"^^xs:double;
21  cs:y3 "-0.076172"^^xs:double;
22  cs:z3 "-0.000000"^^xs:double.
23  _:bond31_1 cs:bondedAtom1 _:atom31_1;
24  cs:bondedAtom2 _:atom31_2;
25  cs:label "Bond1";
26  cs:bondLength "0.960000"^^xs:double;
27  cs:bondOrder "1"^^xs:integer.
28  _:bond31_2 cs:bondedAtom1 _:atom31_1;
29  cs:bondedAtom2 _:atom31_3;
30  cs:label "Bond2";
31  cs:bondLength "0.960000"^^xs:double;
32  cs:bondOrder "1"^^xs:integer.
```

CHEMICAL SEMANTICS

Ontologies

Wikipedia - An **ontology** formally represents knowledge as a set of concepts within a domain, and the relationships between pairs of concepts. It can be used to model a domain and support reasoning about concepts.

Domain – computational chemistry
Reasoning – about properties of molecules

Web Ontology Language (OWL)
e.g. Chemical Entities of Biological Interest (ChEBI) database plus ontology
e.g. Protege (Stanford) – software for creating and manipulating ontologies

Ontologies

ChEBI – Chemical Entities of Biological Interest

– Public Domain

Project of **EMBL-EBI**

European Bioinformatics Institute (Cambridge) of European Molecular Biology Lab (Heidelberg)

Database of 29,000 molecules

Ontology for molecular characterization e.g.

Methane is_a alkane

Methane is_conjugate_acid_of methanide

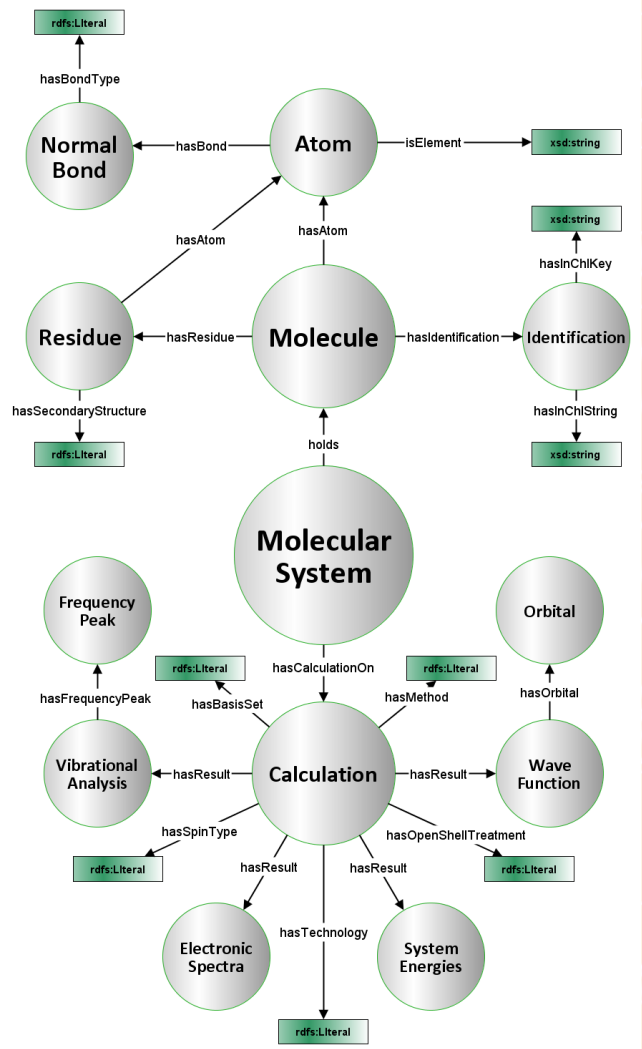
CHEMICAL SEMANTICS

Chemical Semantics Ontology

Gainesville Core

http://purl.org/gc/

gc.owl

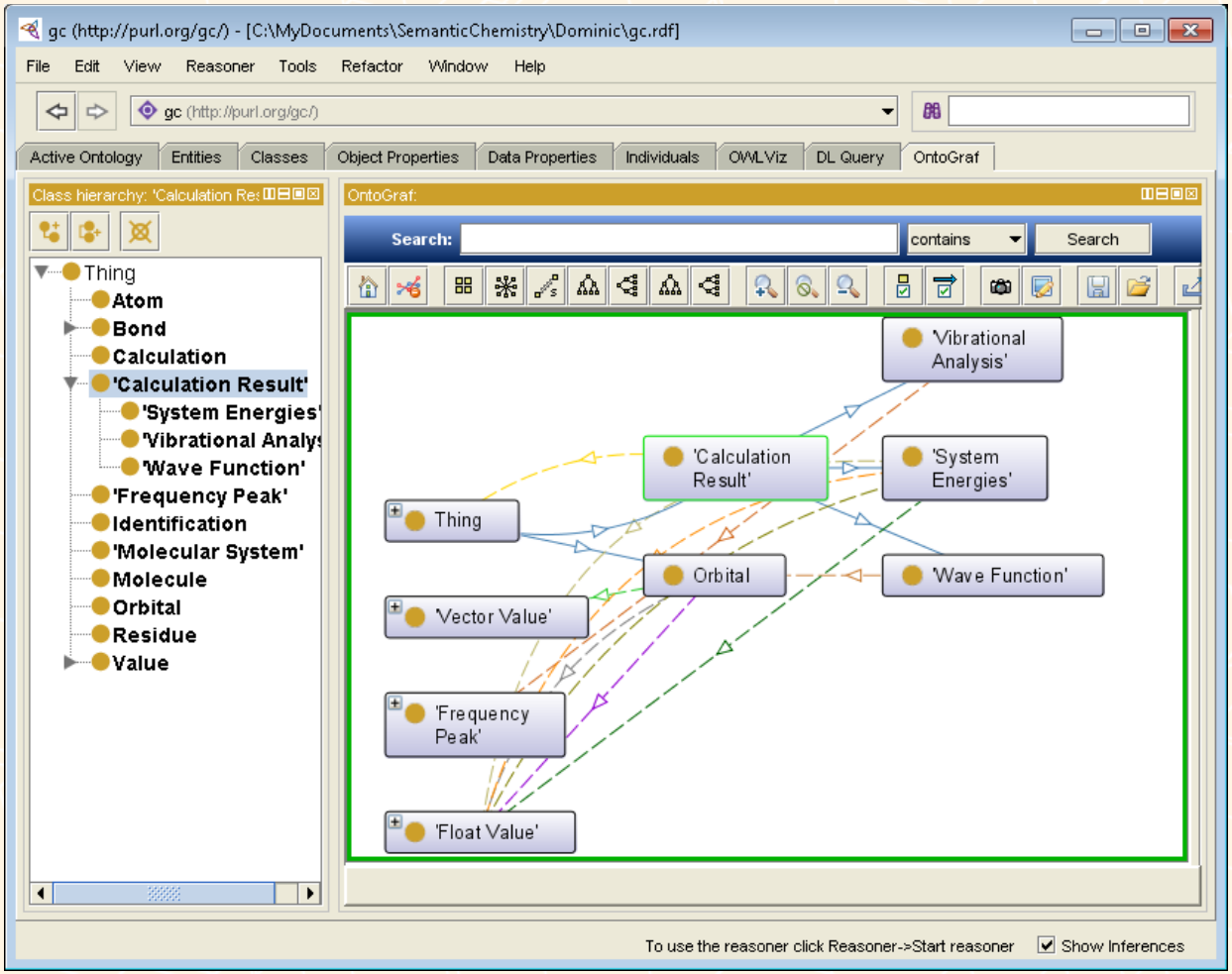


MakoLab | Hypercube

CHEMICAL SEMANTICS

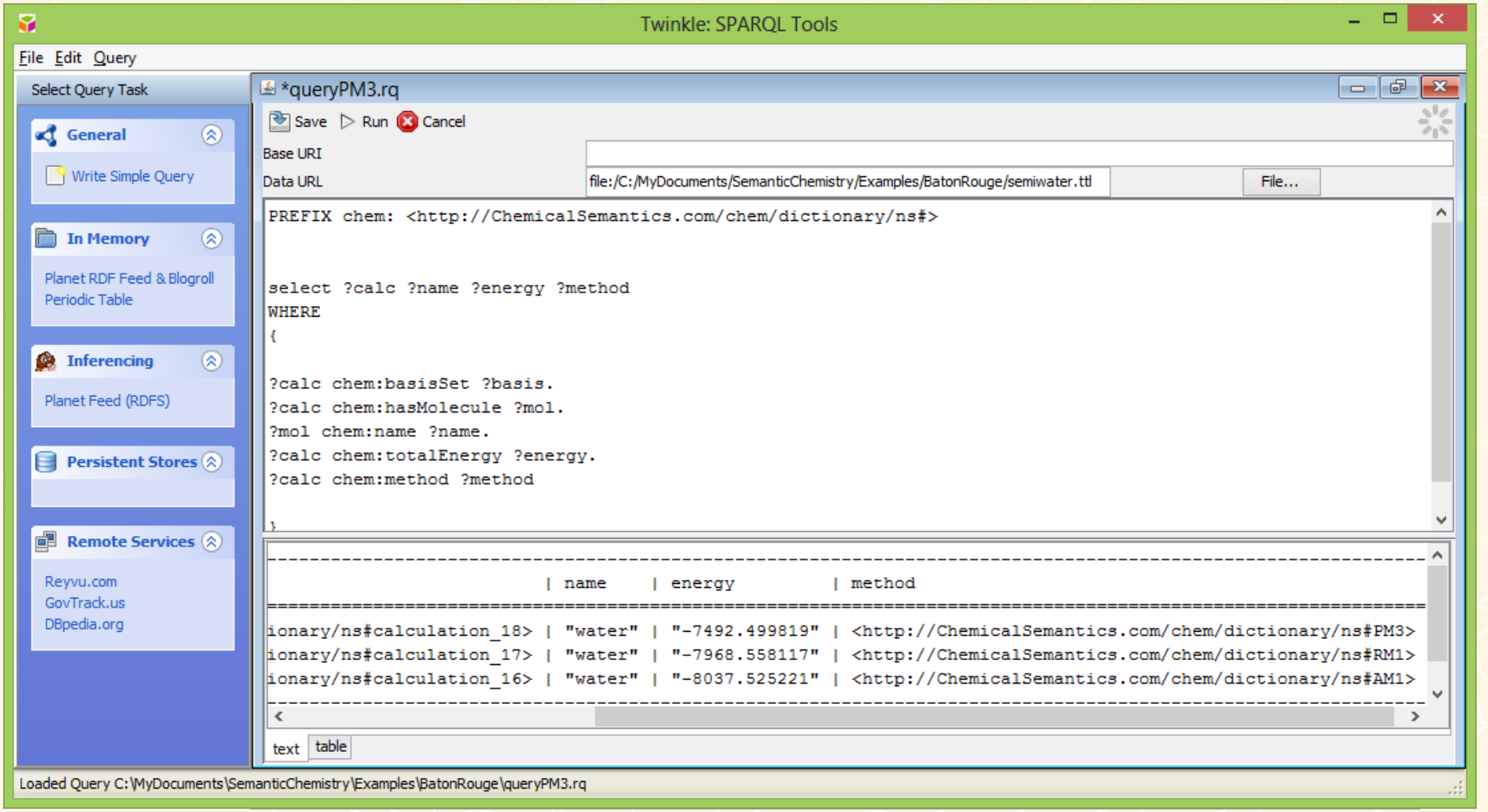
Chemical Semantics Ontology

gc.owl with Protege



CHEMICAL SEMANTICS

SPARQL Query



MakoLab | Hypercube

CHEMICAL SEMANTICS

Conclusion

An emerging successor to the web, **the Semantic Web**, will likely profoundly change the very nature of how scientific knowledge is produced and shared, in ways that we can now barely imagine.

CHEMICAL SEMANTICS

Thank you...

Mirek Sopek
MakoLab SA

Demokratyczna 46, 93-430 Lodz, Poland

Phone: +48 600 814 537
Web: www.makolab.com
eMail: sopek@makolab.com

Neil Ostlund,
Hypercube, Inc.

1115 NW 4th St. Gainesville, FL 32608, USA

Phone: (352) 371 7744
Web: www.hyper.com
eMail: ostlund@hyper.com