Applying Ontologies And Semantic Web Technologies To Environmental Sciences And Engineering

Master's Thesis Defense

Candidate: Viral Parekh

Advisors
Dr. Jin-Ping (Jack) Gwo
Dr. Timothy Finin

May 6, 2005

Outline

- Introduction
  - Problem Description
  - Approach
  - Use Case Applications
  - Motivation
- Related Work
- Ontology Development Process
  - Technologies
  - Methodology

Outline

- Ontologies
  - Environmental Ontology
  - Molecule Ontology
  - Metadata Ontology
  - Models Ontology
- Applications
- Discussion
- Conclusion

Problem Description

- Environmental Sciences and Engineering
  - Complexity and diversity of domain knowledge

- Large volumes of data available
  - Different formats, schemas and semantics
  - Data interoperability problems
  - Difficulty in data discovery and data integration

- Vital need for domain semantics

Approach

- Use of Semantic Web technologies and Ontologies
  - Common framework to allow data sharing and reuse
  - Machine understandable semantics
  - Shared domain models
- Development of domain ontologies
  - Describe domain knowledge
  - Provide semantic metadata for datasets and domain models
  - Efficient mechanisms for data discovery, data interoperability and knowledge sharing

Use Case Applications

- Case 1: A research scientist wishing to model groundwater contamination
  - Acquire the knowledge of models, gather and analyze data, transform data and perform modeling
  - *Semantic descriptions of models and datasets can automate this task*
  - *Composition of sequence of model runs possible*

Use Case Applications

- Case 2: Engineers needing information to conduct preliminary studies
  - Gather and analyze varieties of data
  - *Knowledge base of semantic metadata for datasets can automate this task*
  - *Ontology based searches possible*

Use Case Applications

- Case 3: A Geochemist wanting to study the behavior of different molecules
  - Gather data about molecules and search for geochemical model
  - *Standard semantic knowledge of chemical molecules and reactions can automate the entire task*

Motivation

- Environmental systems demand semantics
- Ontologies provide shared, common vocabulary and domain semantic knowledge
  - Interoperability among heterogeneous datasets
  - Conceptual schema for any dataset
  - Content based discovery and retrieval
  - Semantic descriptions for environmental models
  - Use of standard languages like RDF and OWL
  - Reuse for multiple applications
  - Reasoning and inferencing power

Related Work

- USGS FGDC metadata
  - Text based complex syntactic metadata
- GeoSemantic Web
  - Geographic ontologies for geospatial applications
  - Integration of geographic information with other information
- Earth Systems Grid
  - Discovery and secure access to datasets
  - Ontologies to describe the datasets

Related Work

- SWEET (Semantic Web for Earth and Environmental Terminology)
  - Ontologies and semantic framework for earth sciences
  - Ontology aided search tool
- Hydrologic ontologies and tools for hydrologic datasets
  - Based upon FGDC Metadata standards
- Ontology based system for earthquake sciences

Ontology Development Process

- Technologies
- Methodology

Technologies

- RDF (Resource Description Framework)
  - To describe and relate resources
  - Flexible graph based model
  - Unordered collection of triples
  - Resources identified by unique URIs
- RDFS (RDF Schema)
  - Class definitions and relationships
  - Property definitions and association with classes

Technologies

- OWL (Web Ontology Language)
  - Extensive vocabulary and more expressive
  - Designed for ontology descriptions
  - 3 variants with increasing levels of complexity and expressiveness
    - OWL Lite
    - OWL DL
    - OWL Full

Technologies

- Protégé Ontology Editor
  - Widely used GUI editor for ontology development
  - OWL plugin and ezOWL plugin
- Jena
  - Widely used Java framework for Semantic Web applications
  - Rich API for RDF, RDFS and OWL
  - RDQL to query and retrieve data from knowledge base
  - Persistence for RDF models through backend relational database (MySQL)

Methodology

Process of Ontology development:

1. Defining the domain concepts as classes in the ontology
2. Determining the relationships among these concepts/classes
3. Defining the properties of the concepts/classes
4. Determining the domain and range of the defined properties
5. Defining various class level and property level restrictions if required
6. Finally, creating the knowledge base by identifying the various instances of the defined concepts

Based on Ontology Development Guide 101

Methodology

- Glossaries/Dictionaries
  - USGS, EPA, FGDC, ORNL ESD
- Online libraries of ontologies
  - schemaweb, protégé library
- Interactions with domain expert
- Combination of top-down and bottom-up development process

Methodology

- Formulation of a set of questions
  - Define the scope of ontologies
  - Determine range of applications that could benefit
- Overall Goal
  - Semantic interoperability among heterogeneous datasets

Methodology

Questions

- What is the exact geographic location of this environmental entity or environmental instrument?
- Is rock a type of porous medium? Is Basalt a type of igneous rock?
- What are the rainfall measurements for this Rain Gauge during the month of March 2005?
- What are the possible attributes and the different types of Soil?

Methodology

Questions

- Can we perform geochemical modeling on the chemical species present in the groundwater in this well located in Baltimore, MD? If yes, how?

- What are the chemical species found inside this sample of water? Do these chemicals react to form a particular compound, if not what are the possible outcomes?

- What are the types of Computational Models available in order to perform analyses of the climate data to predict weather patterns?

Molecule Ontology **Models Ontology**

Methodology

Questions

- What is the temporal and spatial extent for this dataset?
- Give me all the identification information for this dataset.
- How do I retrieve and use this dataset?
- What type of information does this dataset contain?
- What is the format of this dataset?
- Can we track the provenance for this dataset in order to determine the trust level?

Ontologies

- Environmental Ontology
- Molecule Ontology
- Metadata Ontology
- Models Ontology

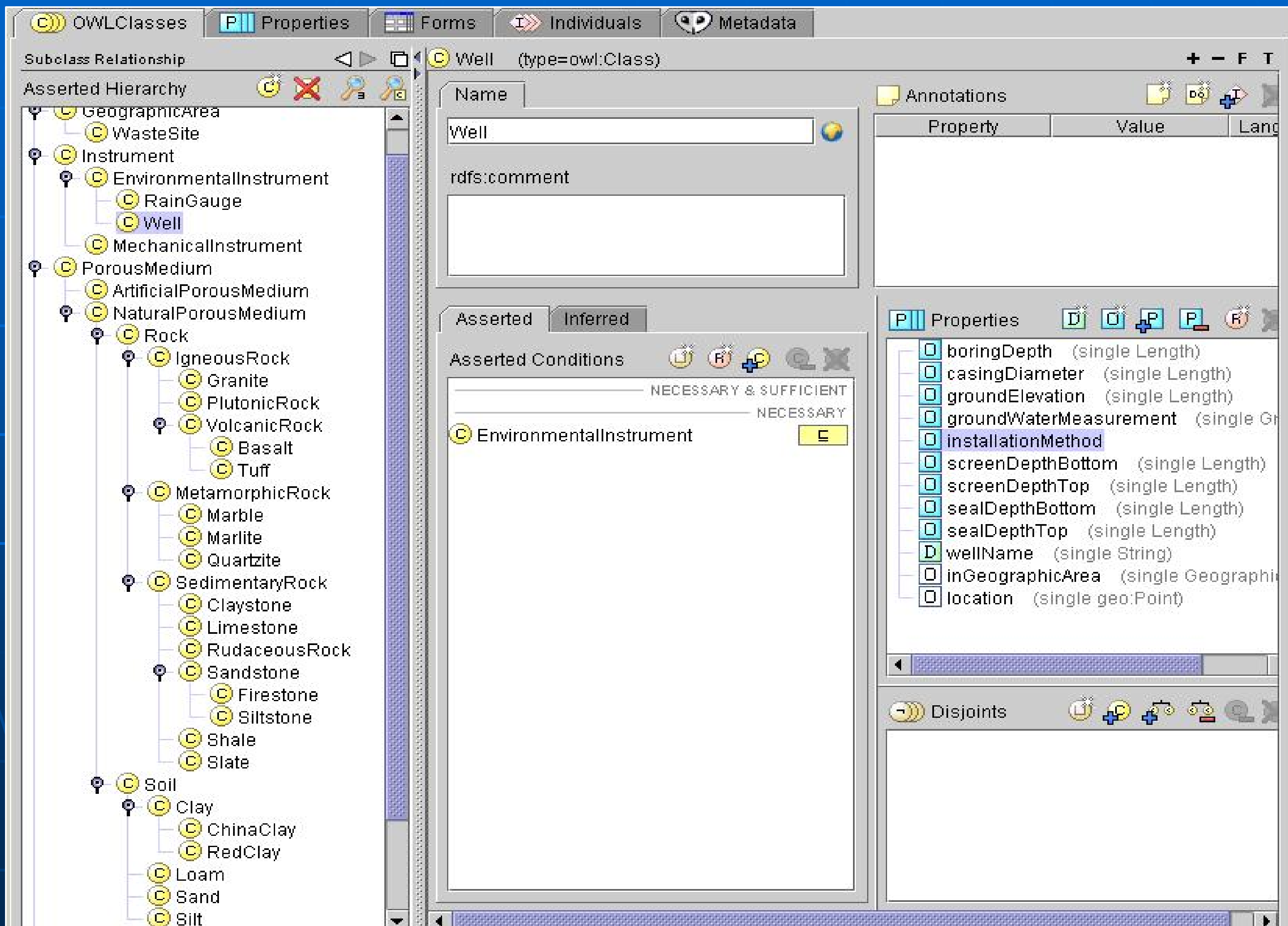
Environmental Ontology

- Domain knowledge through description of concepts like Rainfall, Groundwater, River, Rock, Soil, etc and related properties

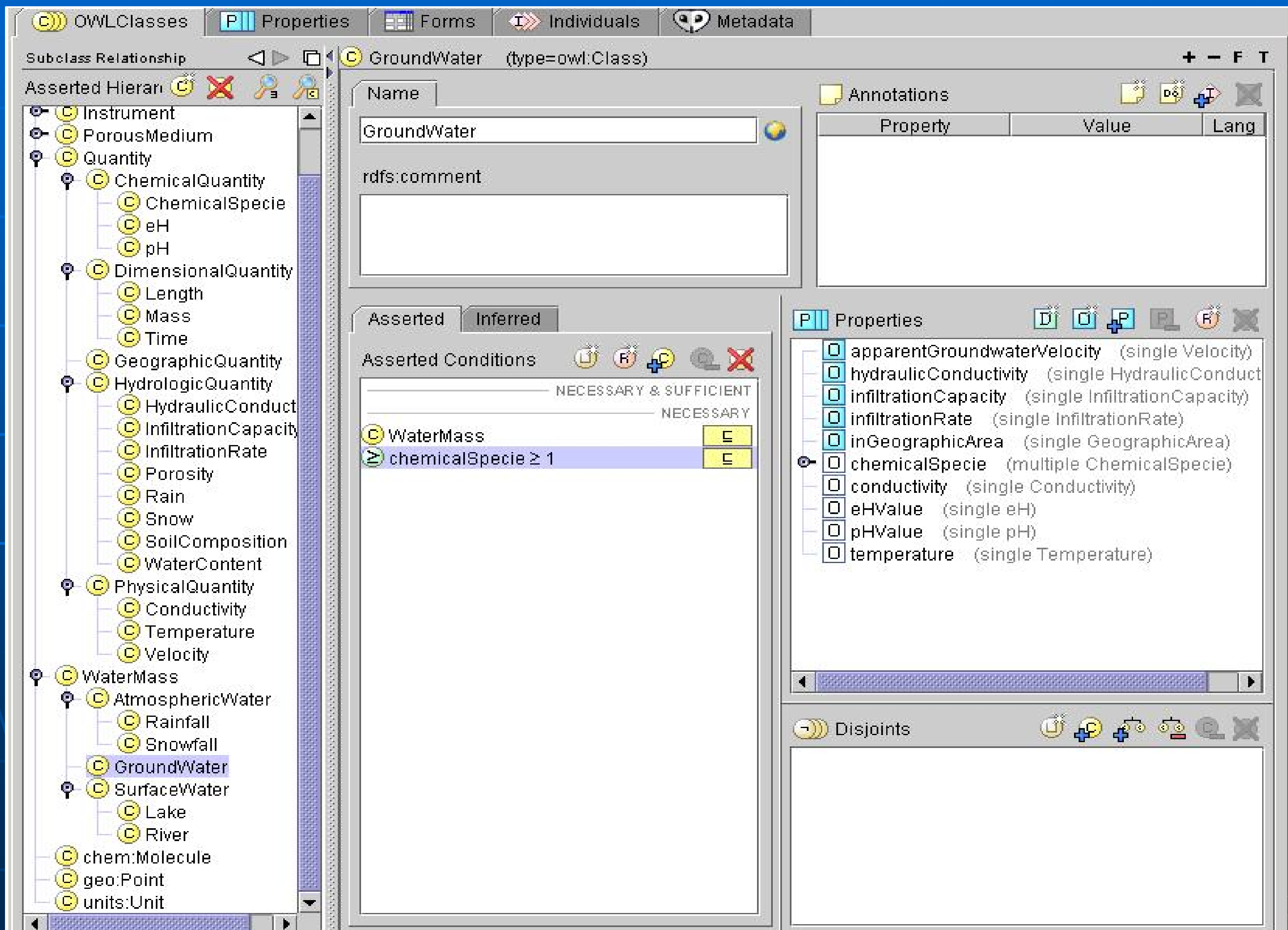
- Definitions of different environmental instruments like Rain Gauge, Well, etc

- Provision of recording measurements

Environmental Ontology



Environmental Ontology



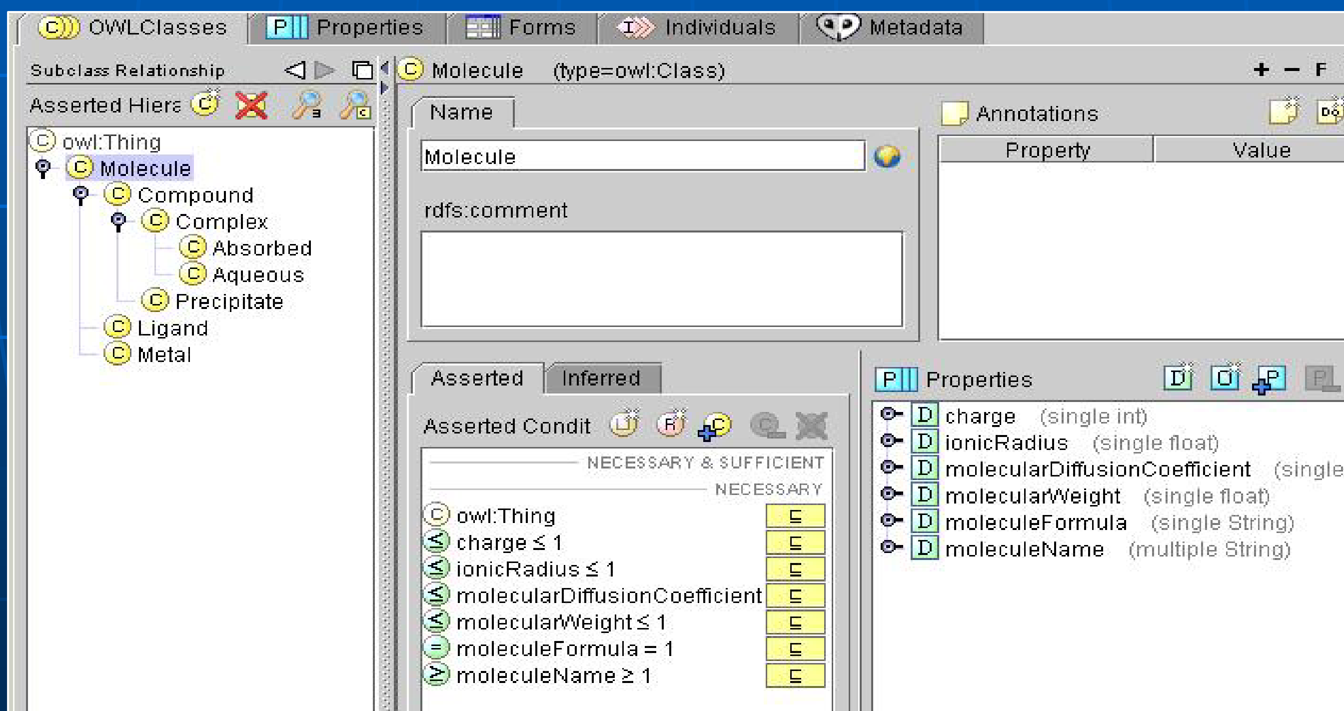
Environmental Ontology

- Geographic Ontology
  - Minimalistic RDF vocabulary which describes Points with latitude, longitude and altitude
  - RDFIG Geo vocab workspace http://www.w3.org/2003/01/geo/

- Units Ontology
  - Part of SWEET ontologies
  - Several characterizing classes are defined such as *Unit, BaseUnit, DerivedUnit, UnitDerivedByRaisingToPower, SimpleUnit, ComplexUnit, Prefix, UnitDerivedByScaling, PrefixOrUnit, UnitDerivedByShifting, etc*
  - Includes definition of units such as meter, minute, hour, degree, Newton, kilogram_meterSquare_perSecondSquare, volt, pascal_perSecond, coulomb, etc

Molecule Ontology

- Provides a knowledge base of all kinds of chemical molecules and their properties

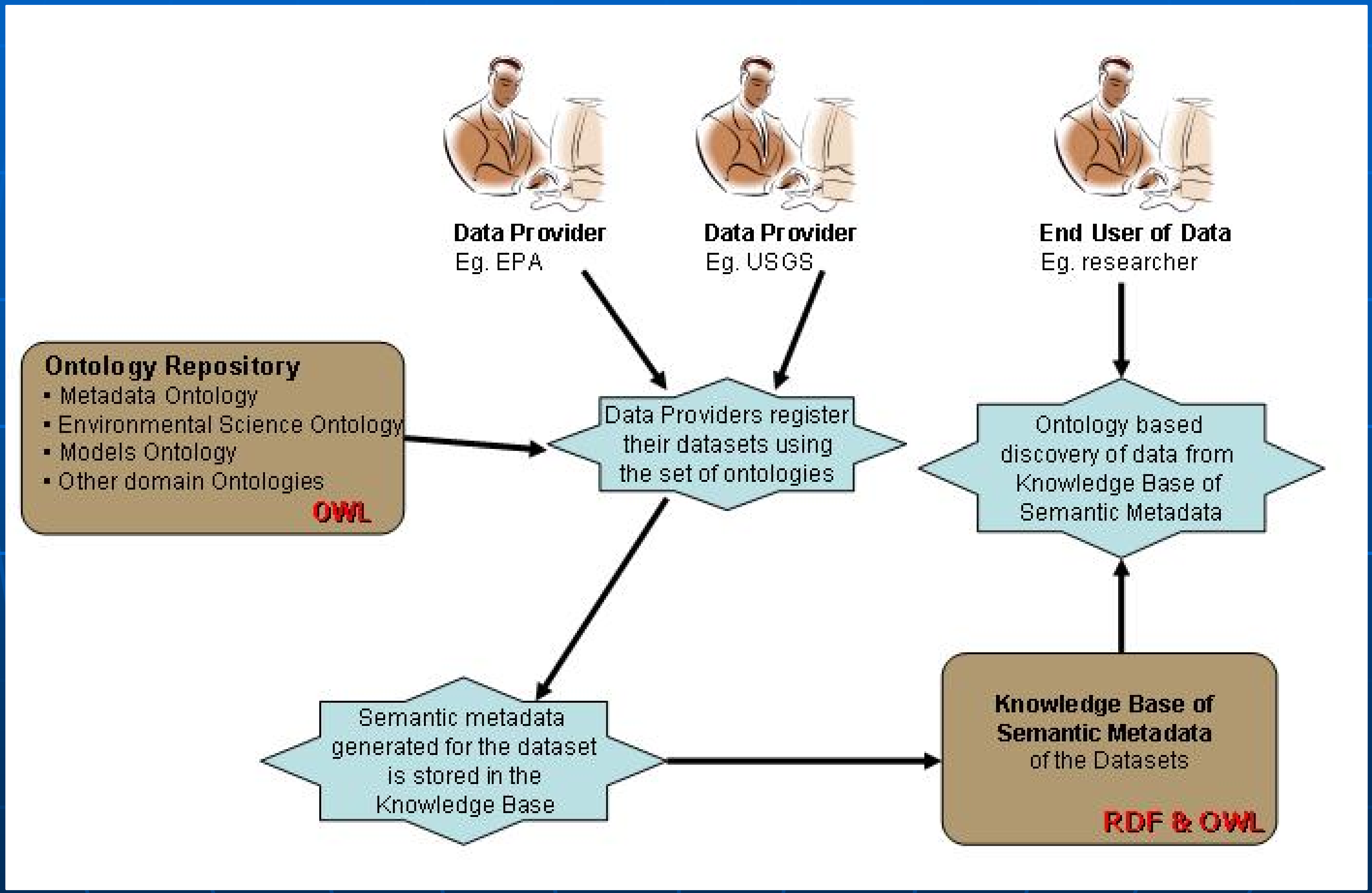


Metadata Ontology

- Provides meta-information and semantic description for environmental datasets
- Generates a conceptual schema for the dataset
- Goal: content based search and retrieval of data

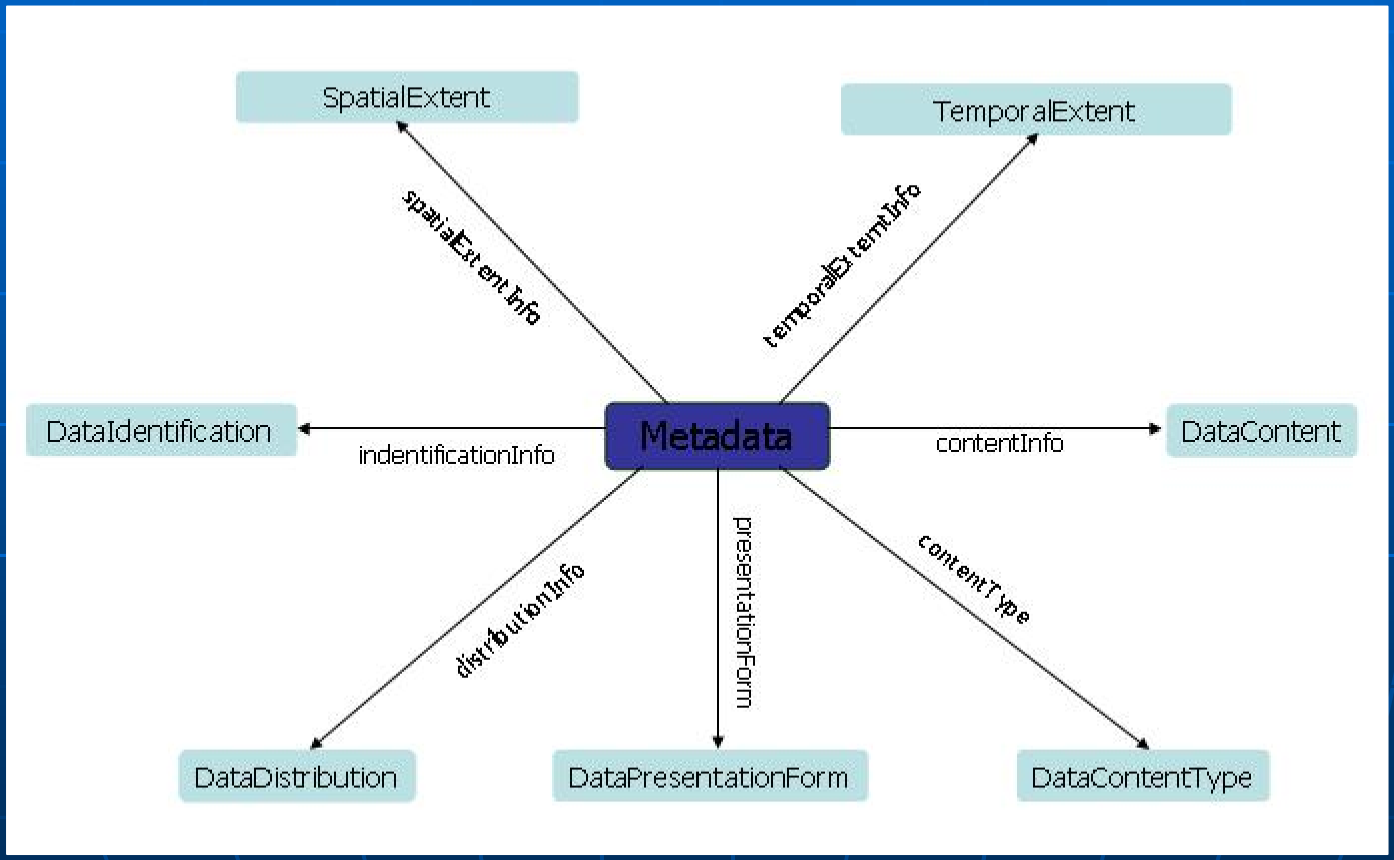
V. Parekh, J. Gwo and T. Finin, **“Ontology based Semantic Metadata for Geoscience Data”**, Proceedings of The 2004 International Conference of Information and Knowledge Engineering

Metadata Ontology



Role of Metadata Ontology

Metadata Ontology



Ontology elements

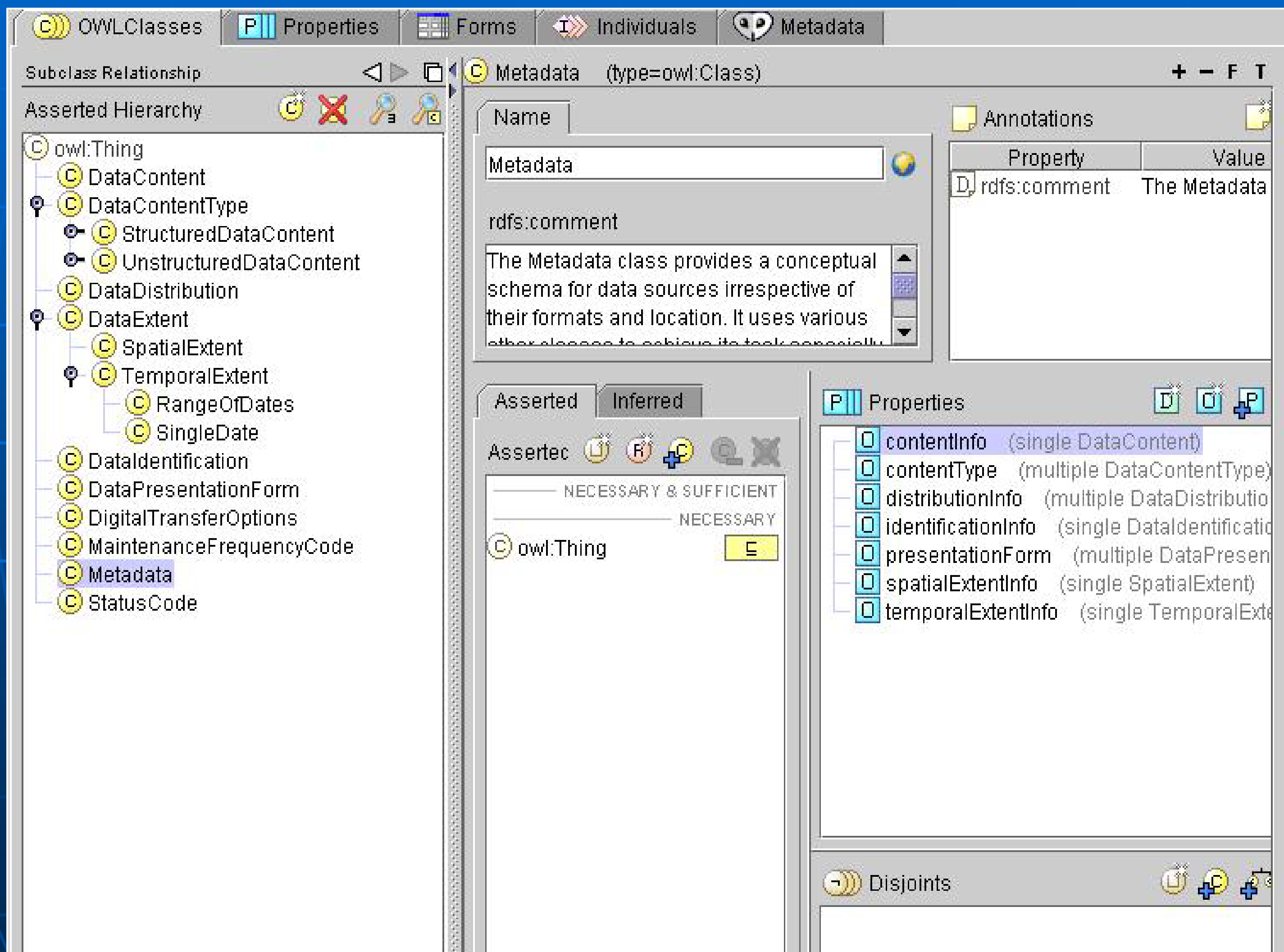
Metadata Ontology

- DataIdentification
  - *title, description, publication, note*
  - *creator, participant, pointOfContact*
  - *creationDate, lastModificationDate*
  - *status, maintenanceFrequency*
  - *isPartOf, isDerivedFrom*
- SpatialExtent
  - *eastBoundLongitude, northBoundLatitude, southBoundLatitude and westBoundLongitude*
- TemporalExtent
  - *beginDate, endDate* and just *date*

Metadata Ontology

- DataContent
  - *hasConcept* and *hasRelation*
  - Links back to domain ontologies
- DataContentType
  - Indicates whether *StructuredDataContent* or *UnstructuredDataContent*
- DataPresentationForm
  - Indicates whether *digital* or *hardCopy*
- DataDistribution
  - *accessConstraints, distributionFormat, distributor, legalDisclaimer, transferOptions* and *useConstraints*

Metadata Ontology

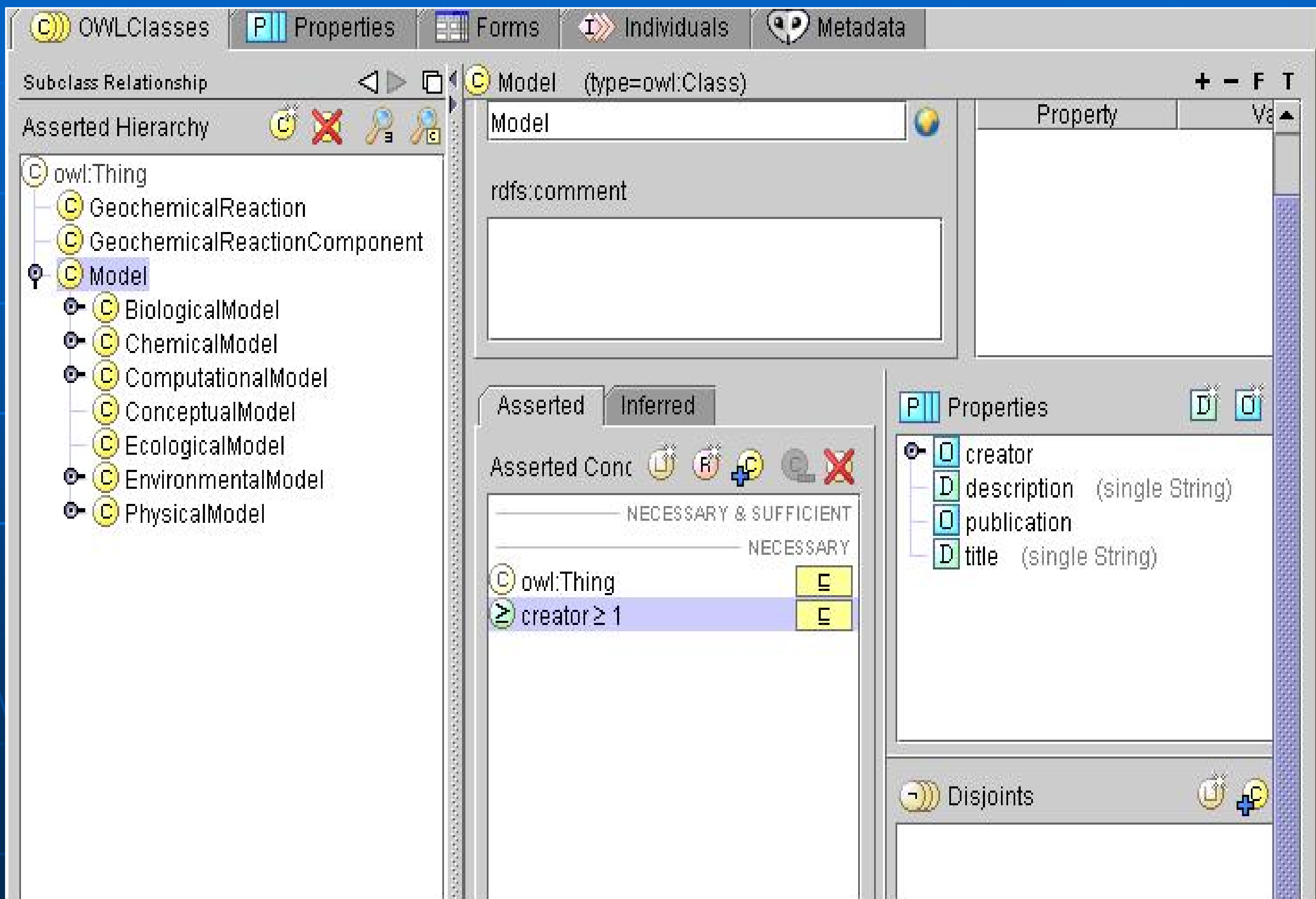


Models Ontology

- Definition and description of various domain models and tools
  - Biological, Physical, Computational, Chemical, Environmental, Ecological, etc

- Provide model run descriptions, identification of input data, model configuration and documentation

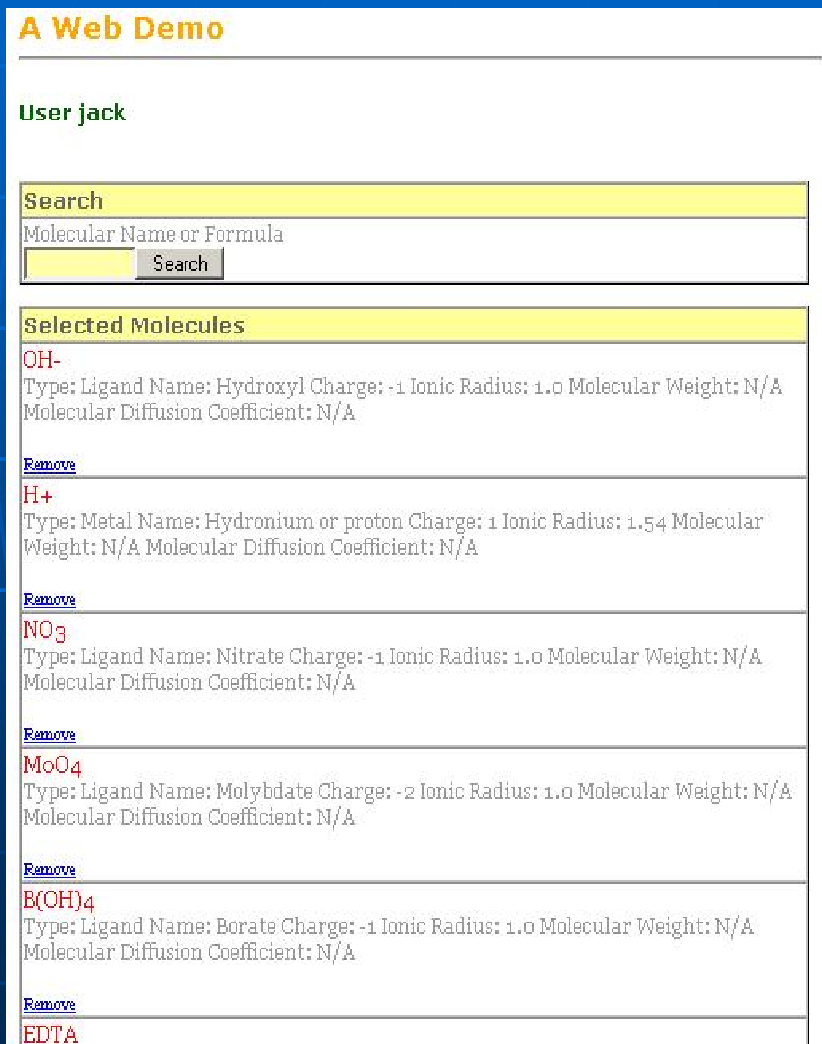
Models Ontology

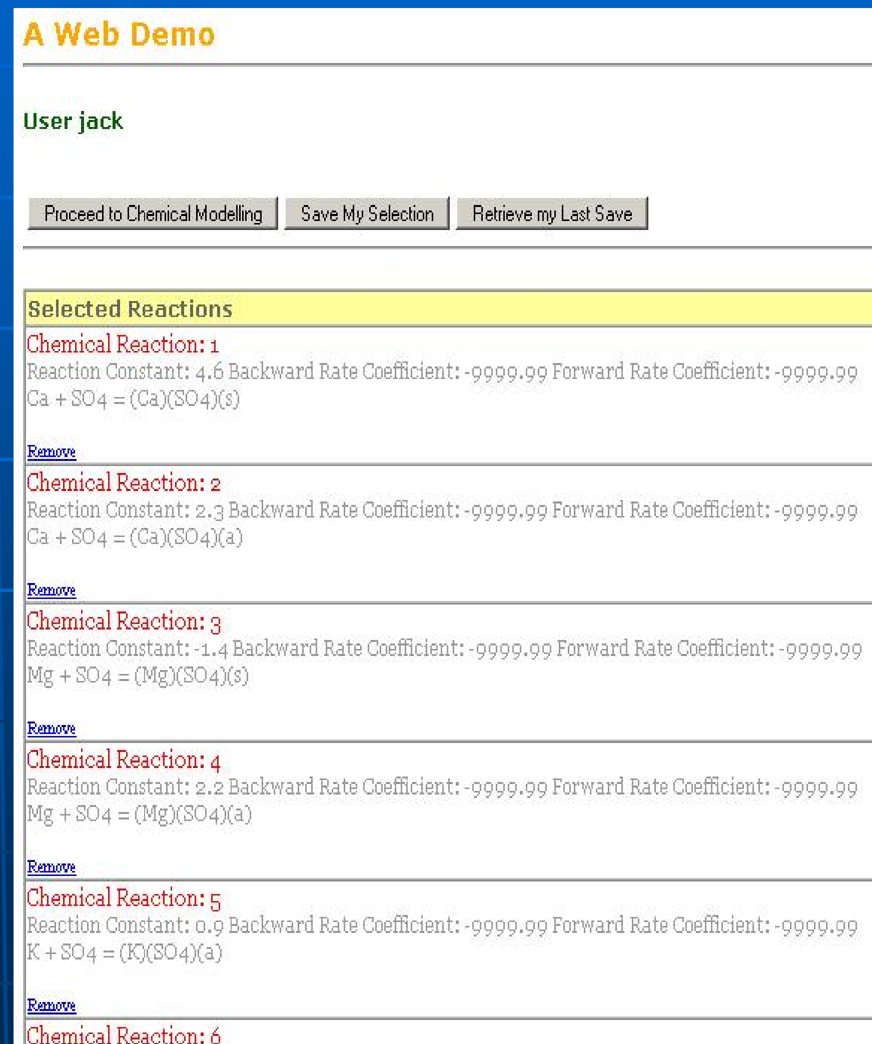


Applications

- 2 typical applications in the geochemical and groundwater hydrology communities
- Application 1: geochemist wanting to do modeling of chemical species for soil samples
  - Use of Molecule and Models ontologies and knowledge base
  - Process
    - Search and select molecules
    - Retrieve the chemical reactions
    - Search and select the geochemical model
    - Run the model

Applications





Applications

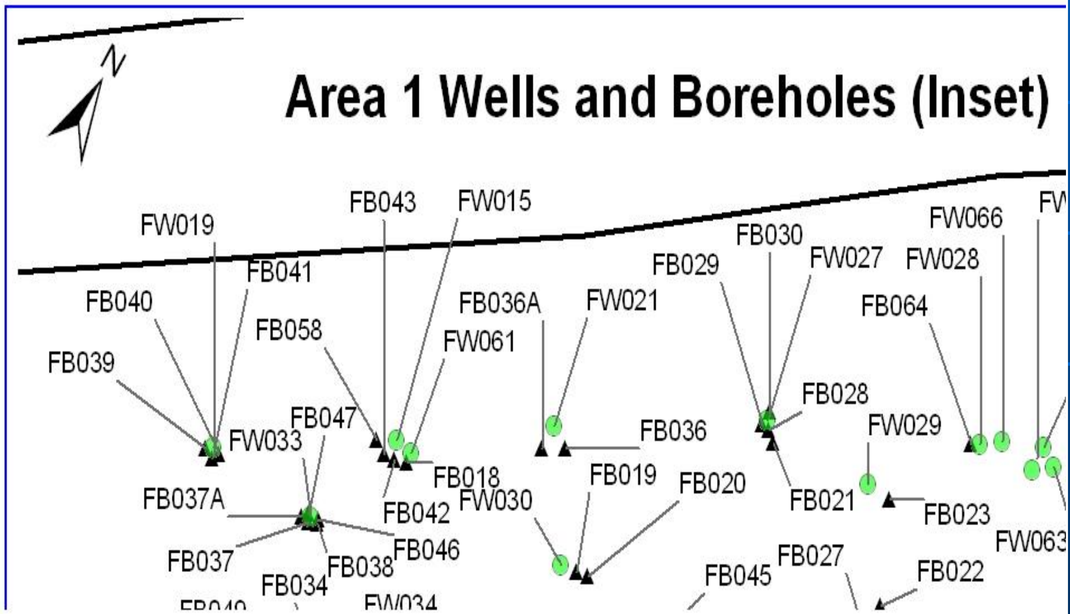
- Application 2: A geochemist wants to do study distributions of chemical pollutants in the wells of a waste site
  - Use of Environmental, Molecule and Models ontologies
  - Process
    - View and select any well from the waste site
    - View semantic metadata including the chemical species knowledge for the selected well
    - Use chemical modeling knowledge base to retrieve chemical reactions
    - Search and select geochemical model
    - Run the model

Applications

Ontological Framework for Environmental Systems

A Web Demo

Tip- The following wells have data: ED06 ED07 EU06 EU07 FW011 FW012 FW013 FW014 FW015 FW016 FW017 FW018 FW019 FW020 FW021 FW022 FW023 FW027 FW028 FW
FW030 FW031 FW032 FW033 FW034 FW062 FW063 FW064 FW065 FW066 GW-127 GW-245 GW-246 GW-247 GW-276 TPB10 TPB11 TPB19 TPB31 TPB32 WD3 WU3



Well Specification

Name: FW011
Latitude: 35.976658
Longitude: 84.27314
GroundElevation: 1004.95
BoringDepth: 24.67
CasingDiameter: 1.05
ScreenDepthTop: 20.84
ScreenDepthBottom: 23.69
SealDepthTop: 9999.99
SealDepthBottom: 9999.99

Groundwater Physical Properties

Conductivity 1.406
Temperature 19.6
eH Value -307.0
pH Value 6.18

Groundwater Chemical Properties

Molecule: Br Concentration: 0.35
Molecule: Cl Concentration: 50.1
Molecule: NO3 Concentration: 19.58
Molecule: SO4 Concentration: 104.5
Molecule: U+4 Concentration: 0.0733

Proceed to Chemical Modelling using Geochem program

Discussion

- More complex and realistic applications need to be demonstrated
- Ontology standardization efforts needed by bodies such as EPA, USGS and NASA
- Better URI naming required

Discussion

- Automated/Semi-Automated tools needed for faster ontology development
  - Use of dictionaries/glossaries and domain text
  - Statistical text mining techniques
  - Machine learning strategies

Conclusion

- Information infrastructures for efficient data sharing and integration
  - Ontologies and Semantic Web technologies like RDF and OWL
- Intelligent environmental information systems
  - Efficient data discovery mechanisms
  - Planning and execution of models
  - Effective decision making and resolution of imminent environmental problems