

# Vocabulary/Semantics Standardisation in Earth Sciences

Necessary for Representation of Interoperable Knowledge in Computers

## **Three Semantics-Based Systems developed by GOL** **which require Logically Consistent\* Taxonomies**

Minerals Exploration



Metallurgy



Landslide Hazard Mapping



A presentation by  
Clinton Smyth

[cpsmyth@msn.com](mailto:cpsmyth@msn.com)

\* In the strict sense of "Logic as used for Human or Computer Reasoning"

# The OneGeology Project

A Global-Scale Project which also needs Logically Consistent Taxonomies

The screenshot shows the OneGeology website homepage. At the top left is the OneGeology logo, which consists of a globe with a red and green landmass and the text 'ONE Geology'. To the right of the logo is the tagline 'Making Geological Map Data for the Earth Accessible'. Below the tagline is a navigation bar with links for 'Home' and various languages: العربية, 中国, English, Français, Русский, and Español. On the left side, there is a vertical menu with expandable sections: 'What is OneGeology', 'Participants', 'Organisation', 'Getting involved', 'Technical overview', 'Technical detail for participants', 'Meetings', 'Portal', 'OneGeology eXtra', and 'Press information'. The main content area features a 'Welcome to OneGeology' section with a globe made of puzzle pieces and text describing the project's launch in 2007 and its goal of creating dynamic geological map data. To the right of this is a section titled 'New Zealand and Japanese earthquake and tsunami' with flags of Japan and New Zealand, and text expressing support from the OneGeology community. Below that is a 'Monthly news' section with an image of a rolled-up map. At the bottom left, there is a 'NEW extended content to the Onegeology kids pages' section with a video player showing a map of Earth Processes. To the right of this is a section titled 'The 6th Technical Working' with a 'New publication' section showing a geological map. The bottom right corner of the website features a cartoon illustration of various geological features like mountains, rocks, and fossils, each with a character name like George, Maps, Rocks and minerals, Fossils and dinosaurs, Earth, Eddi, Anthony, Larry, Manuel, Paltra, and Gordon.

OneGeology subscribes to GeoSciML and its Controlled Vocabularies

# A 2011/2012 Geoscience BC Project

United States



```

classDiagram
    class EarthMaterial {
        id
        name
        parent
        position
    }
    class ConstituentPart {
        id
        position
        part
    }
    class OrganicMaterial {
        name
        chemicalFormula
    }
    class InorganicFluid {
        name
        chemicalFormula
    }
    class Mineral {
        name
        chemicalFormula
    }
    class CompoundMaterial {
        name
        chemicalFormula
        mineralCategory
        mineralSubcategory
    }
    class RockMaterial {
        name
        mineralCategory
        mineralSubcategory
    }
    class PhysicalDescription {
        id
        name
        description
        capability
        laboratory
        category
    }
    class FabricDescription {
        name
        type
    }
    class MineralogyProperties {
        name
        mineralogy
    }
    class MineralogyDescription {
        name
        mineralogy
    }

    EarthMaterial "1" -- "0..1" ConstituentPart
    EarthMaterial "1" -- "0..1" OrganicMaterial
    EarthMaterial "1" -- "0..1" InorganicFluid
    EarthMaterial "1" -- "0..1" Mineral
    EarthMaterial "1" -- "0..1" CompoundMaterial
    CompoundMaterial "1" -- "0..1" RockMaterial
    RockMaterial "1" -- "1" PhysicalDescription
    PhysicalDescription "1" -- "1" EarthMaterial
    RockMaterial "1" -- "0..1" FabricDescription
    RockMaterial "1" -- "0..1" MineralogyProperties
    MineralogyProperties "1" -- "0..1" MineralogyDescription
    MineralogyDescription "1" -- "0..1" EarthMaterial
  
```

- DEPARTMENT OF  
PRIMARY INDUSTRIES
- EarthResourceML Background
- Information Model for Exchange
  - Australian Mines Atlas
- Scenarios
  - deliver data to Australian Mines Atlas
  - mineralisation investigation over an exploration lease
  - land use planning to aid land use decisions.
  - a mine site hazard study
  - examine mineral system potential
- Governance
  - Australian Government Geologists Information Committee (GGIC)
  - Mineral Occurrence Task Group
  - Extension of GeoSciML
- Demonstration

Australia (Federal and State agencies)

All these are moving towards GeoSciML and its Controlled Vocabularies



# Vocabulary Standardisation

## A Critical Pillar to Computational Intelligence

- Justification: Knowledge Interoperability
  - Standard Vocabulary is required for expressing human knowledge in computers
  - Standard Vocabulary is required for computer communication to other agents
- Resources Required
  - Vocabulary Repositories
  - Software Tools: Ontology and Taxonomy Editors; Mediators
  - Multi-Disciplinary Staffing Capacity
- April 2012 Status
  - “Conventional” Applications showcasing the need for Standards
  - Standards Development and Deployment
  - Available Software Tools
  - “Computational Intelligence” Applications using Standards?



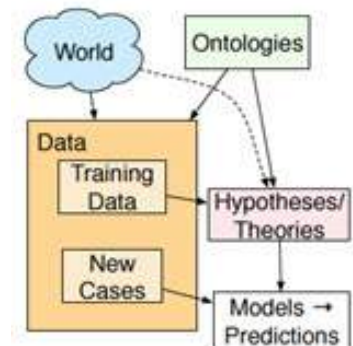
# Computational Intelligence requires Knowledge Interoperability

Knowledge Interoperability requires that we use Standard Vocabularies

- General Problem Statement

- We need a system that can record theories about, and learn theories from, our data in order to make (probabilistic) predictions

[David Poole, [Towards a Logic of Feature-Based Semantic Science Theories](#), Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010)]



- Example Interoperable Knowledge Input Problem

- We need to capture and interoperate Mineral Deposit Models from both the USGS and the British Columbia Geological Survey [\[www.georeferenceonline.com/minematch/\]](http://www.georeferenceonline.com/minematch/)

- Example Interoperable Knowledge Output Problem

- We need to output mineral identification data from a quantitative XRD analyser directly into an agricultural fertility mapping system (or a metallurgical plant optimisation system)



# Resources Required to Enable Knowledge Interoperability

## The Imperatives of Semantic e-Science and the Semantic Web

### ■ Vocabulary Repositories

- Global repositories are available for reference;  
Local repositories need to be harvested and mediated



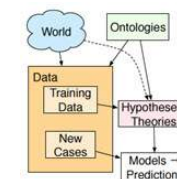
### ■ Software Tools

- Mediators  
Ontology Editors  
Taxonomy Editors



### ■ Multi-Disciplinary Human Resources

- Earth Science, Computer Science, Logic  
Probability, Semantics, Human-Computer Interfaces



### ■ Promotional Applications

- Without these, we cannot attract funding  
See: [www.OneGeology.org](http://www.OneGeology.org)



# Global Vocabulary Repositories: GeoSciML

Enabling Global Interoperability while permitting Local Specialisation

The screenshot shows the SEE GRID community website. The header includes the SEE GRID logo, the text "SEE GRID community website", the CSIRO logo, and navigation links for "Jump" and "Search". The main content area is titled "CGIModel" and includes a breadcrumb trail: "You are here: SEEGrid > CGIModel Web > InteroperabilityWG > ConceptDefinitionsTG (12 Jul 2011, SteveRichard)". There are links for "Log In or Register", "Edit", and "Attach". The main heading is "CGI Geoscience Concept Definitions working group". Below this, there are several links organized into categories: "Task Group Membership" (Reorganization of CGI semantic interoperability work groups), "Where to find CGI vocabularies" (HTML formatted views of 201001 version vocabularies), "Hot topics" (New vocabulary requirements), "Notes" (201012 vocabulary update tags posted, Adoption of http URI, OWL Ontology for Lithology, Subversion repository rearranged), and "Version 200811 Vocabularies" (Links to individual version 200811 vocabularies). On the right, there is a "See also" section with links to various working groups: Interoperability working group, Use-cases and Requirements task group, GeoSciML Design task group, Service Architecture task group, Implementation Test Bed task group, Technical Assistance task group, Geoscience Concept Definitions task group, and International Geoscience Portal? working group. At the bottom right, there is a diagram showing a hierarchical structure of geoscience concepts, including "EarthMaterial", "ComponentMaterial", "RockMaterial", "PhysicalDescription", and "EarthMaterial".


[http://onegeology.org/technical\\_progress/geosciml.html](http://onegeology.org/technical_progress/geosciml.html)



# Local Vocabulary Repositories

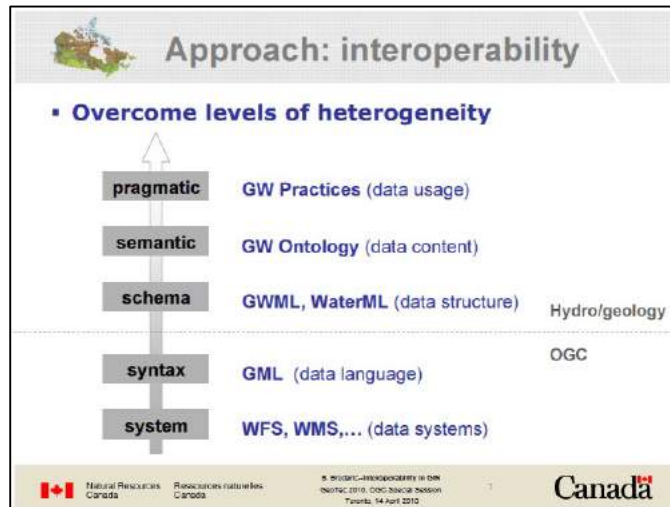
## Maintaining Local Specialisation while enabling Global Interoperation

### European Union

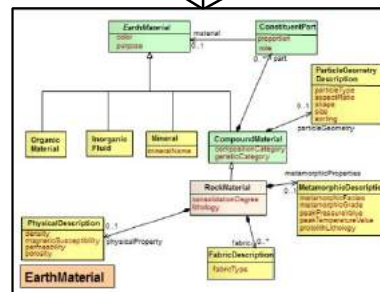
 **INSPIRE**  
Infrastructure for Spatial Information in Europe

**D2.8.III.21 Data Specification on Mineral Resources – Draft Guidelines**

<b>Title</b>	D2.8.III.21 INSPIRE Data Specification on <i>Mineral Resources</i> – Draft Guidelines
<b>Creator</b>	INSPIRE Thematic Working Group <i>Mineral Resources</i>
<b>Date</b>	2011-06-14
<b>Subject</b>	INSPIRE Data Specification for the spatial data theme <i>Mineral Resources</i>
<b>Publisher</b>	INSPIRE Thematic Working Group <i>Mineral Resources</i>



Canada



### United States

 **GEOSPATIAL PLATFORM**

WELCOME TO THE GEOSPATIAL PLATFORM WEBSITE

The partner agencies of the Federal Geographic Data Committee (FGDC) are developing a Geospatial Platform to more effectively provide place-based products and services to the geospatial public. The Geospatial Platform will be a managed portfolio of common geospatial data, services, and applications coordinated and administered by national, regional, and local agencies and stakeholders. The Geospatial Platform is a shared infrastructure for use by government agencies and partners to meet their common needs and the broader needs of the Nation.

**Key Feature:** *Imagery The Response to the 2011 Earthquake in Japan*

COMMON DATA	COMMON SERVICES	COMMON APPLICATIONS
		

**DEPARTMENT OF PRIMARY INDUSTRIES** **EarthResourceML Background**

- Information Model for Exchange**
  - Australian Mines Atlas
- Scenarios**
  - deliver data to Australian Mines Atlas
  - mineralisation investigation over an exploration lease
  - land use planning to aid land use decisions
  - a mine site hazard study
  - examine mineral system potential
- Governance**
  - Australian Government Geologists Information Committee (GGIC)
  - Mineral Occurrence Task Group
  - Extension of GeoSciML
- Demonstration**

Australia (Federal and State agencies)

Vocabulary “Mediators” are the Key

[http://en.wikipedia.org/wiki/Mediator\\_pattern](http://en.wikipedia.org/wiki/Mediator_pattern)

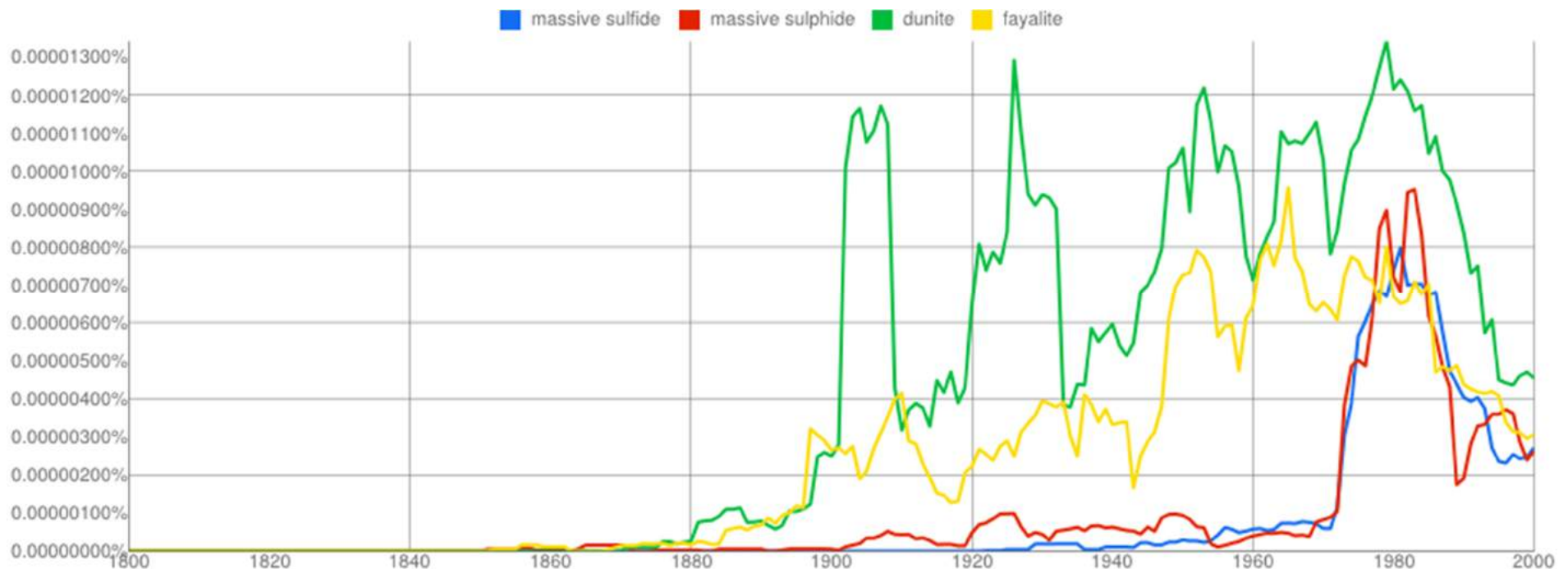
[www.georeferenceonline.com](http://www.georeferenceonline.com)



## Local Vocabulary Repositories: Harvesting/Mediating Outcomes

Need to (constantly) revise the GeoSciML Taxonomy of Rocks

Graph below is a Google NGRAM (word-frequency over time) for  
Massive Sulfide, Massive Sulphide, Dunite, and Fayalite



“Massive Sulfide” is not present in GeoSciML vocabulary, is critical to expression of mineral deposits knowledge, and is present in literature twice as many times as “fayalite”, which is in GeoSciML.

## Vocabulary Software Tools: Ontology Editors

GeoSciML SimpleLithology201001 displayed in TopBraid Composer

Composer infers this hierarchy from these attributes by referring to these hierarchies.

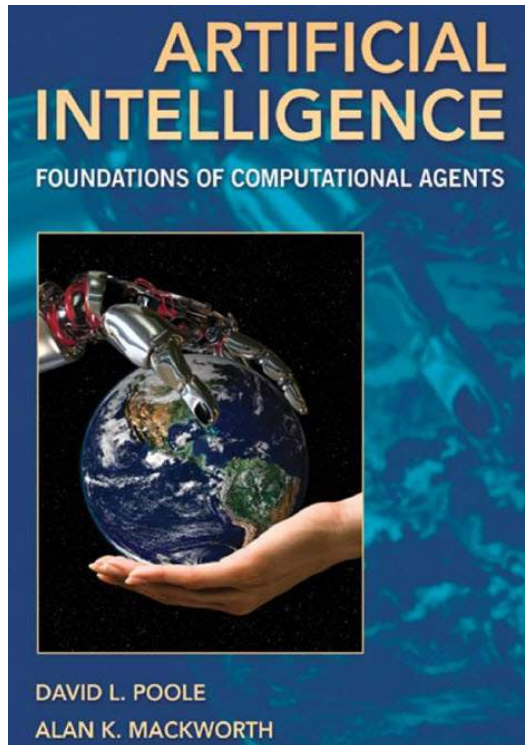
The screenshot shows the TopBraid Composer interface with the following components:

- Navigator:** A tree view on the left showing the ontology structure. The 'owl:Thing' class is highlighted.
- Classes:** A list of classes in the center. The 'earthMaterialsAT:EarthMaterial' class is selected.
- Properties:** A list of properties on the right, grouped by Namespace. The 'earthMaterialsAT' namespace is expanded, showing properties like 'ChemCat', 'ConsolidationDegree', 'FabricType', etc.
- Editor:** The central area shows the 'owl:equivalentClass' definition for 'earthMaterialsAT:EarthMaterial'. It includes a complex logical expression involving 'and' and 'some' operators, along with various property restrictions.

Developed for use by Knowledge Engineering Experts  
Not practical for Domain Experts (geologists, mineralogists, etc)

# The Aristotelian Approach to Taxonomy Specification

## A Necessary Pre-requisite for Intelligent Systems



**“Make Properties  
Explicit in Every  
Taxonomy”**

### Aristotelian Definitions

Categorizing objects, the basis for modern ontologies, has a long history. Aristotle (350 B.C.) suggested the definition of a class *C* in terms of:

**Genus:** a superclass of *C*. The plural of genus is genera.

**Differentia:** the properties that make members of the class *C* different from other members of the superclass of *C*.

He anticipated many of the issues that arise in definitions:

*If genera are different and co-ordinate, their differentiae are themselves different in kind. Take as an instance the genus "animal" and the genus "knowledge". "With feet", "two-footed", "winged", "aquatic", are differentiae of "animal"; the species of knowledge are not distinguished by the same differentiae. One species of knowledge does not differ from another in being "two-footed".*

Note that "co-ordinate" here means neither is subordinate to the other.

This methodology does not, in general, give a tree hierarchy of classes.

Objects can be in many classes.

Each class need not have a single most-specific superclass.

However, it is still straightforward to check whether one class is a subclass of another, to check the meaning of a class, and to determine the class that corresponds to a concept in your head.

In rare cases, this results in a tree structure, most famously in the **Linnaean taxonomy** of living things. It seems that the reason this is a tree is because of evolution.

*Trying to force a tree structure in other domains has been much less successful.*



## Vocabulary Software Tools: Taxonomy Editors

ACE: An Open-Source Taxonomy Editor

Developed by GOL and Université Joseph Fourier (Grenoble, France)

# ACE

## Aristotelian Classes Editor



Philippe Genoud

Danielle Ziebelin

Anthony Hombiat

<http://www.similar2.com:8080/ACE-Editor/?ontology=http://similar2.com/ontologies/earthmaterials201001d.owl>

# Vocabulary Software Tools: Taxonomy Editors

GeoSciML SimpleLithology201001 displayed in ACE

ACE infers this hierarchy from these attributes by referring to these hierarchies.

**Taxonomy**

- Thing
  - EarthMaterial
    - anthropogenic\_material
    - breccia
    - composite\_transformed\_genesis\_material
    - exotic\_evaporite
    - igneous\_material
      - acidic\_igneous\_material
      - basic\_igneous\_material
        - basic\_igneous\_rock
          - doleritic\_rock**
          - foid\_bearing\_diorite
          - tephritoid
        - fragmental\_igneous\_material
        - intermediate\_composition\_igneous\_material
        - ultramafic\_igneous\_rock
      - rock
      - sedimentary\_material
      - unconsolidated\_material

**Attributes**

Property	Value
6 inherited attributes	
ChemCat	intermediate
<input checked="" type="checkbox"/> ChemCat	basic
ConsolDegree	consolidated
GrainSize	fineGrained
GrainSize	phaneritic
ParticleType	crystal
1 intrinsic attribute	
GeneticCateg	hypabyssal

**Value hierarchy**

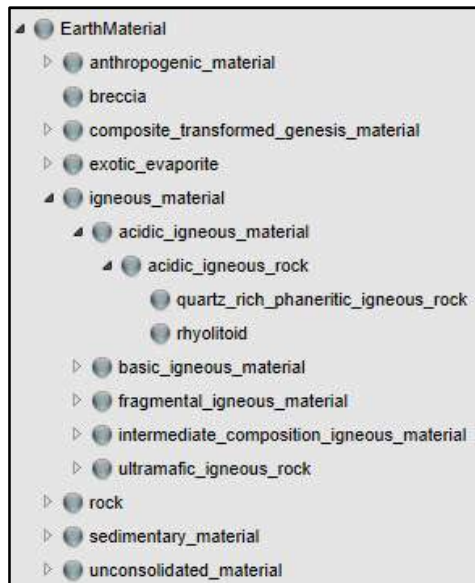
- Thing
  - AnyChemCat
    - acidic
    - aluminous
    - basic**
    - carbonaceous
      - highRank
      - lowRank
      - mediumRank
    - intermediate
    - magnesium\_rich\_igneous\_rock
    - normativeOI\_GT10
    - normativeOI\_LT10
    - siliceous
    - ultraBasic
    - undersaturated

Local ontology loaded. Developed for use by Domain Experts (like geologists) no operation processing

# GeoSciML SimpleLithology201001b: Excel Workflow

“Constituents” column was broken out to 16 independent “dimensions”

## Inferred Taxonomy



	A	B	S	T	U
1	Rock Type	Constituents	ConsolidDegree	MetamorphicGrade	ParticleType
109	amphibolite	amphibolitic	consolidated	-	crystal
110	gabbroic_rock	an_GT_50,M_GT10_LT90,QAPF10	consolidated	-	crystal
111	quartz_gabbro	an_GT_50,M_GT10_LT90,QAPF10asterisk	consolidated	-	crystal
112	foid_bearing_gabbro	an_GT_50,M_GT10_LT90,QAPF10prime	consolidated	-	crystal
113	gabbro	an_GT_50,M_GT10_LT90,QAPF10strict	consolidated	-	crystal
114	alkali_olivine_basalt	an_GT_50,mafic,QAPF9_10	consolidated	-	-
115	foid_monzogabbro	an_GT_50,QAPF13	consolidated	-	crystal
116	foid_gabbroid	an_GT_50,QAPF13_14	consolidated	-	crystal

## 16 New Dimensions

C	D	E	F	G
amph_px_olStatus	apRange	qfRange	AnValue	carbonate
amphibole dominant				
AP_GT50	QF_LT20_GT-10	An>50		
AP_GT50	QF_GT05_LT20	An>50		
AP_GT50	QF_LT-00_GT-10	An>50		
AP_GT50	QF_GT00_LT05	An>50		
AP_GT65	QF_LT20_GT-10	An>50		
AP_GT50_LT90	QF_LT-10_GT-60	An>50		
AP_GT50	QF_LT-10_GT-60	An>50		

Pellet Reasoner

The other 9 dimension were used unchanged as inputs to the Reasoner.

Contact the author if you wish to know more about applying this methodology in Protegé



# Editing SimpleLithology201001 in ACE

Changes made to Properties or their Values immediately update the Taxonomy

Changes made here or here or here immediately modify these hierarchies.

The screenshot shows the ACE interface with the following panels:

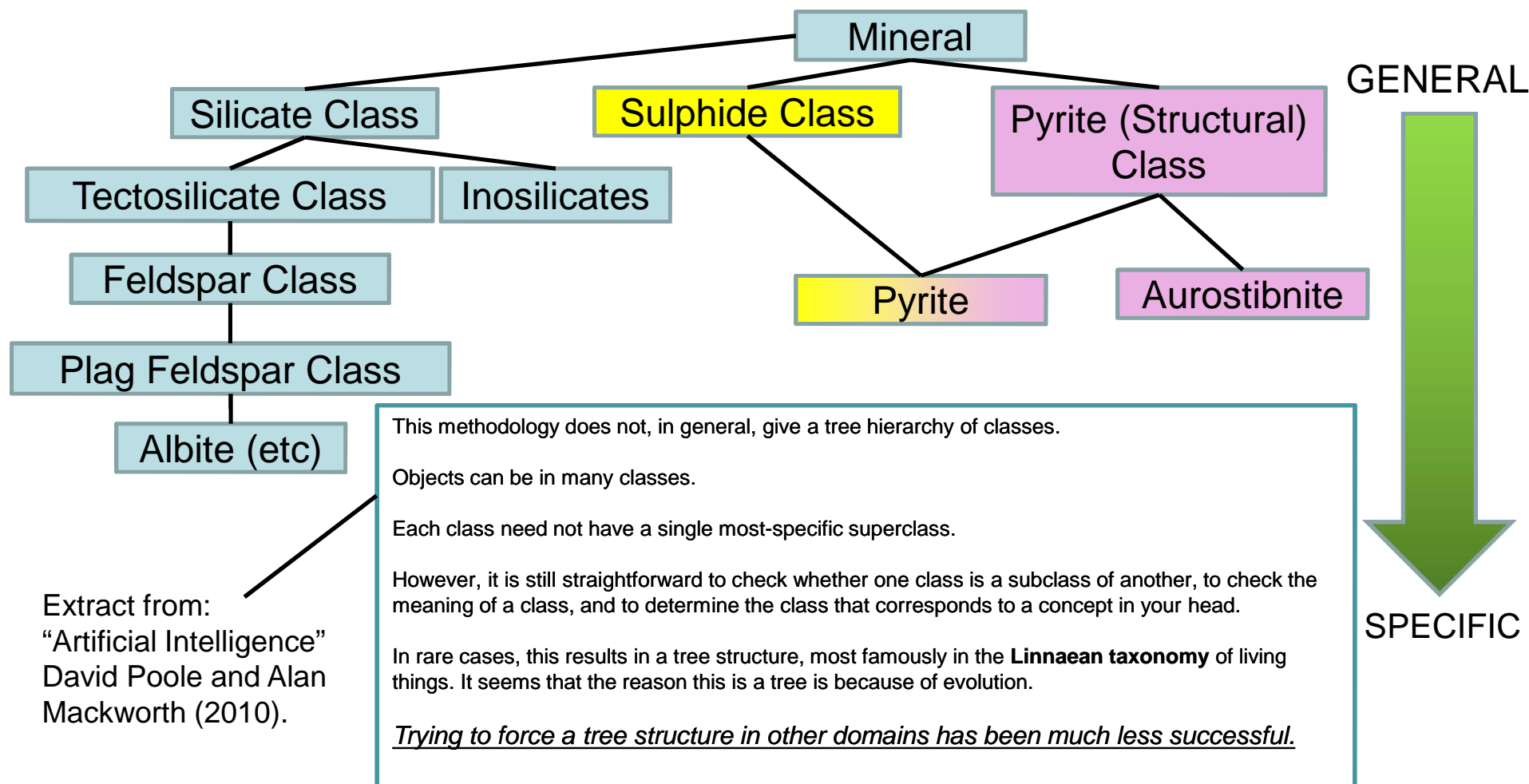
- Taxonomy:** A tree view showing the hierarchy of classes. The 'basic\_igneous\_rock' class is highlighted.
- Attributes:** A table showing the list of attributes for the selected class. The 'ChemCat' attribute is highlighted with the value 'basic'.
- Value hierarchy:** A tree view showing the hierarchy of values for the selected attribute. The 'basic' value is highlighted.

Arrows point from the text 'Changes made here or here or here immediately modify these hierarchies.' to the 'ChemCat' attribute in the Attributes panel, the 'basic' value in the Value hierarchy panel, and the 'basic\_igneous\_rock' class in the Taxonomy panel.

Class and Terminal Node Names in the Taxonomy may appear in multiple locations.

# Minerals Taxonomy Levels: The Aristotelian Solution

Crystal Structure and Composition are Differentia at the Same Level of the Taxonomy

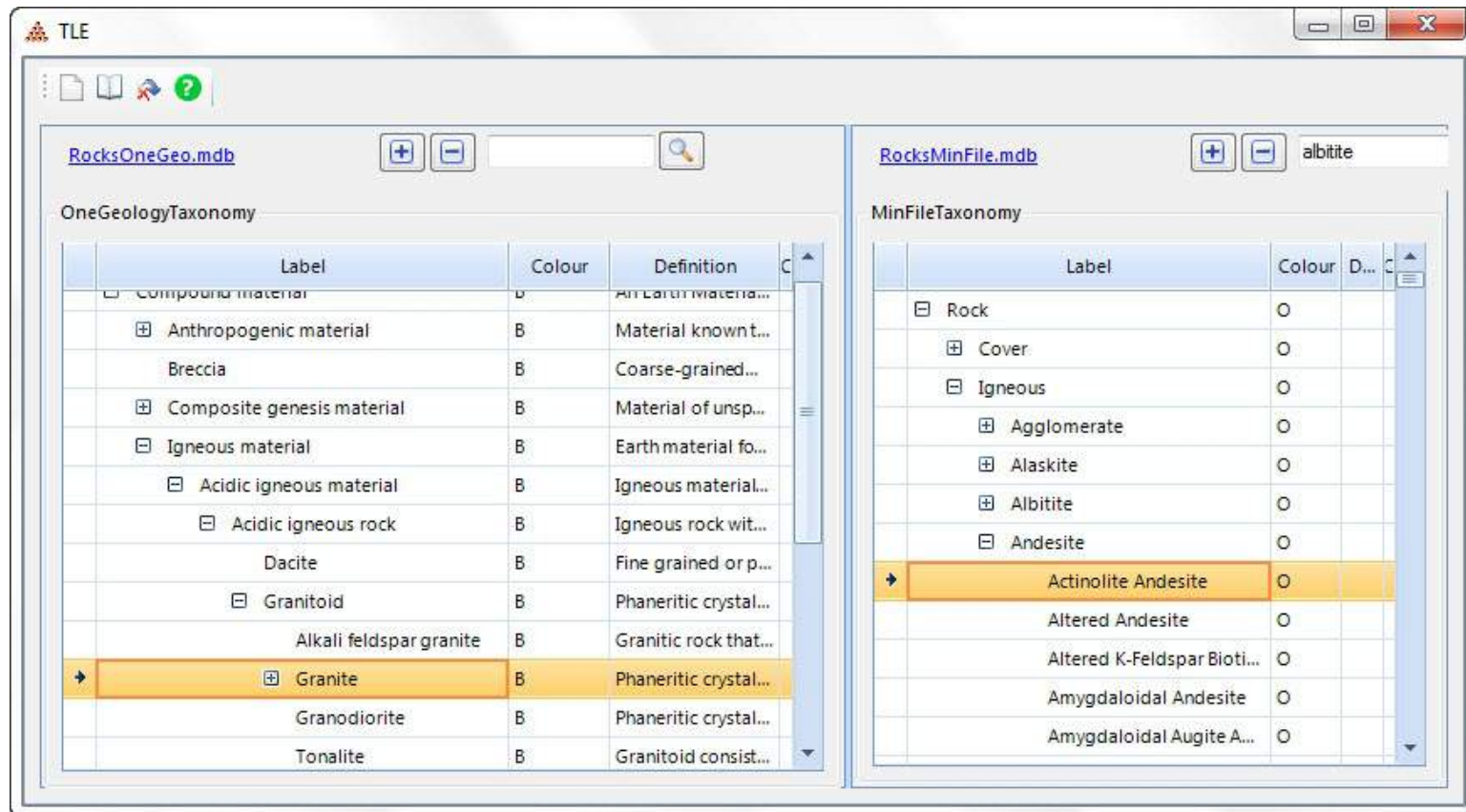


Using Micronex's MinIdent Win4 system, we are creating additional useful “Top Level” compositional classes, such as “Copper Mineral”, “Arsenic-Bearing Mineral”, etc.

# Vocabulary Software Tools: Taxonomy Editors

## TLE: TreeList Editor

Easy-to-Use by Anyone Interested in Nomenclature



Available free from: [www.golinfo.com/tle](http://www.golinfo.com/tle)

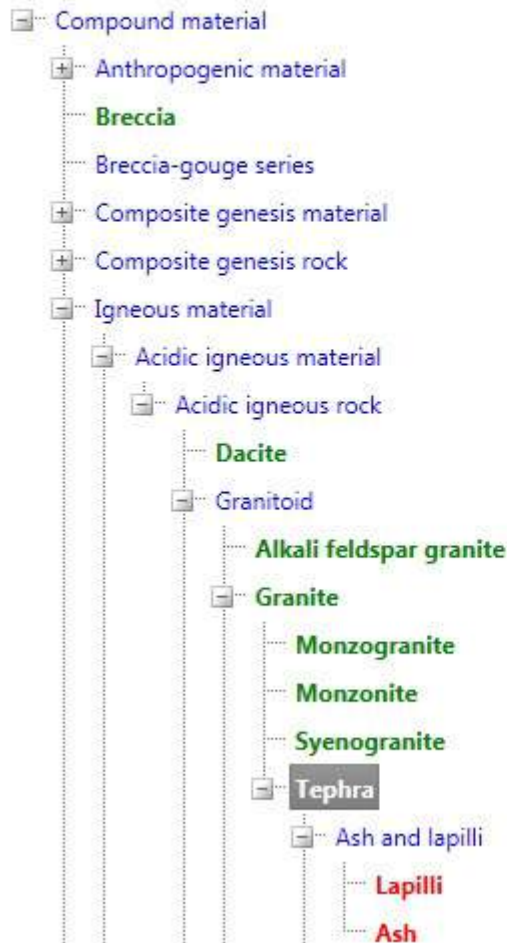


# Vocabulary Software Tools: Taxonomy Editors

## TLE: TreeList Editor Output

### RPDS Common Rock Names mapped to the One Geology Rock Classification

[Home](#)



#### Description

Unconsolidated pyroclastic material in which greater than 75 percent of the fragments are deposited as a direct result of volcanic processes and the deposit has not been reworked by epiclastic processes. Includes ash, lapilli tephra, bomb tephra, block te

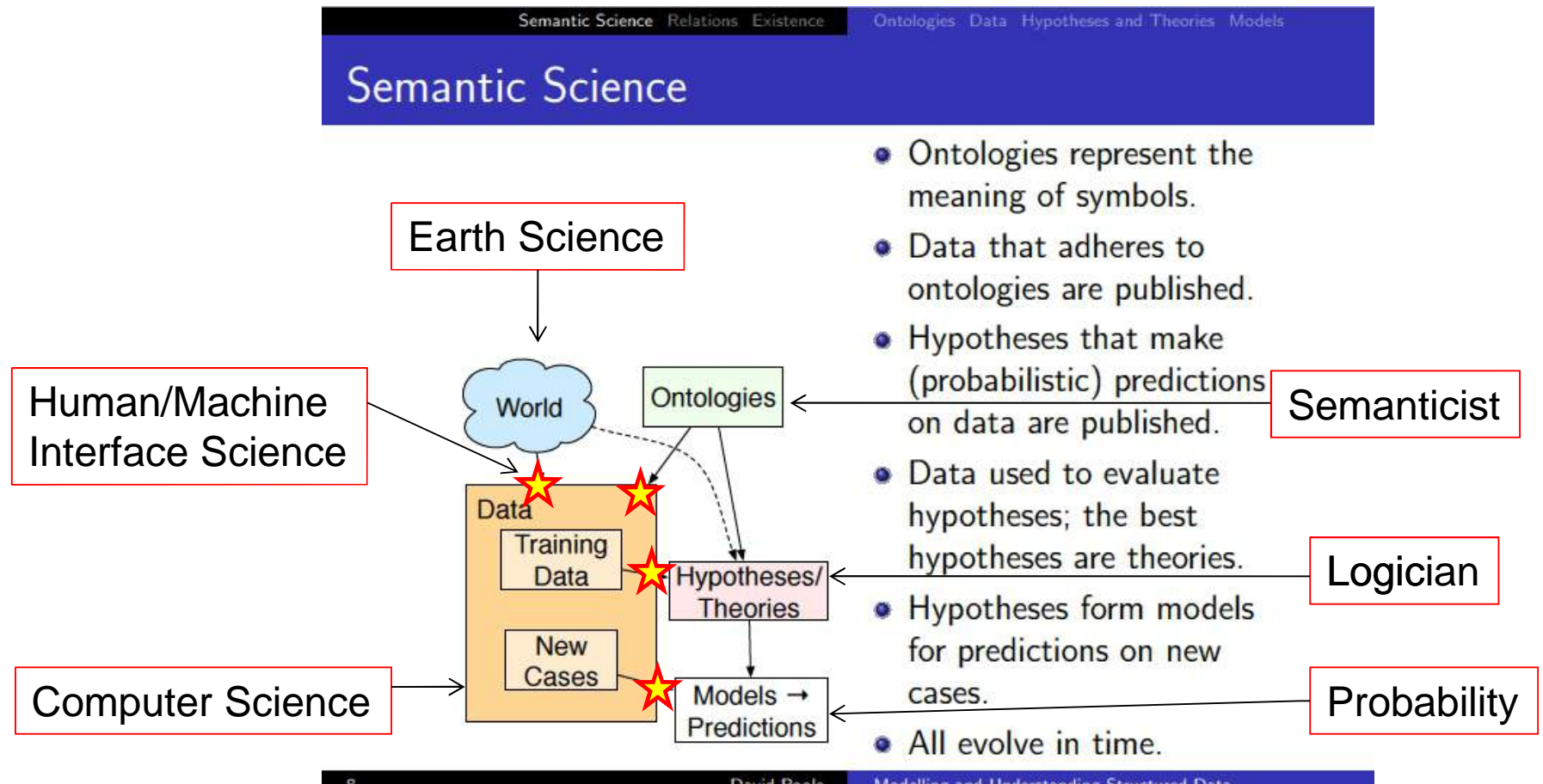
Facilitates Mapping between Taxonomies, and communicating the intricacies and importance of the subject



# Vocabulary Development Human Resources

Must be Multi-Disciplinary

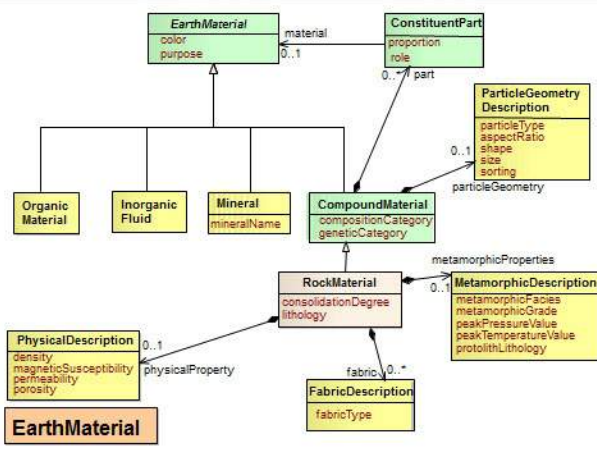
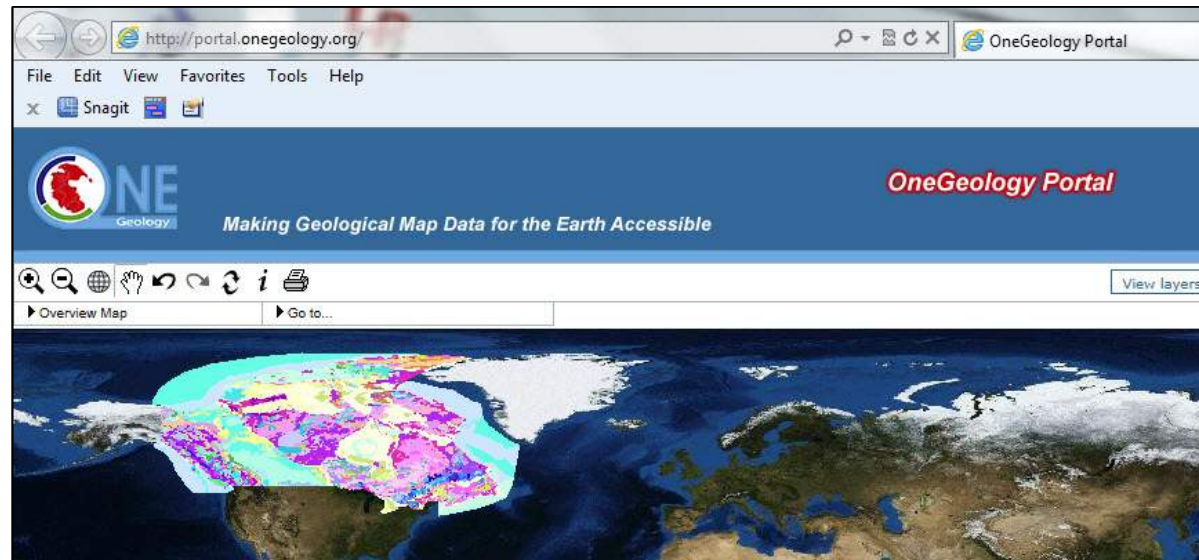
Absence of One Disciplinary Consideration leads to Flawed Vocabulary



Base slide from: David Poole, [Logic, Probability and Computation: Foundations and Issues of Statistical Relational AI](#), invited paper, [11th International Conference on Logic Programming and Nonmonotonic Reasoning](#), Vancouver, May 2011.

# April 2012 Status: OneGeology, INSPIRE and GeoSciML

Setting Standards for the 21<sup>st</sup> Century – using them for better science





# April 2012 Status: GeoSciML adoption in the US

## OneGeology Workshop in Lakewood, CO



[Home](#) > [Diary](#) > [USGIN/OneGeology Workshop](#)

**What is OneGeology** +

**Participants** +

**Organisation** +

**Getting involved**

**Technical overview** +

**Technical detail**

**for participants**

**Meetings** +

**Portal**

**OneGeology eXtra** +

**Press information**



### **USGIN/OneGeology Workshop**

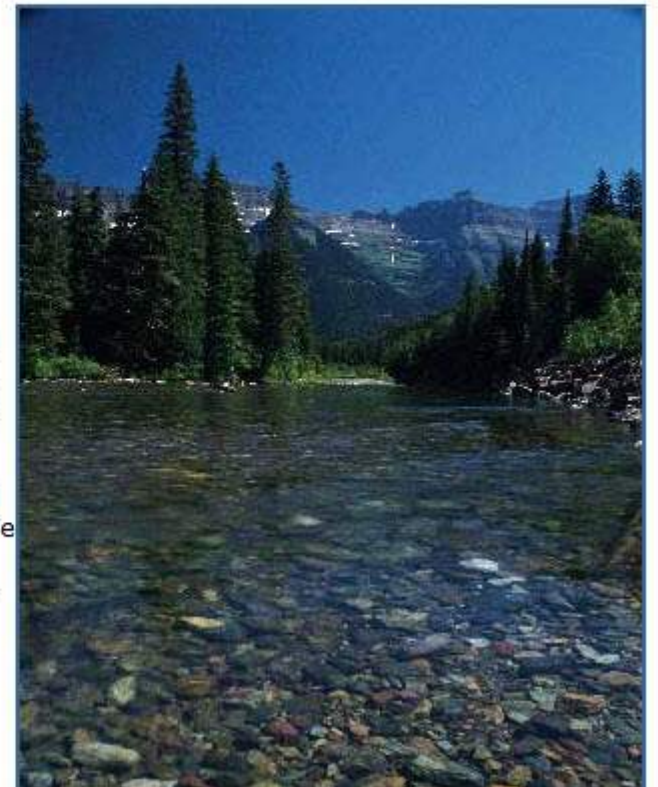
+ Dates: May 3 and 4, 2012 at USGS Ice Core Conference Room, Denver Federal Center, Lakewood, CO

#### **Objective**

+ The goal of this workshop is deploy new GeoSciML portrayal map services as part of a USGIN-sponsored OneGeology-United States collection of services.

#### **Concept**

+ We are looking for participants from the State Geological Surveys and US Geological Survey who have geologic map data sets they are ready to publish as web map and feature services. The data must be in a digital data base format, ready for the necessary extract/transform/load processing to deploy the maps using the OpenGeospatial consortium Web Map service and Web Feature service. We request participants to arrive at the workshop with data and access to a map server (workshop examples will use ArcGIS server and GeoServer, see below on AASG arrangements for service hosting). In advance of the workshop, we will provide participants with tutorials on deploying GeoSciML map and feature services for OneGeology. Workshop sessions will review the service framework and procedure for preparing data, and then



The End



Thank You to Geoscience BC for Financial Support

[www.geosciencebc.com](http://www.geosciencebc.com)