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

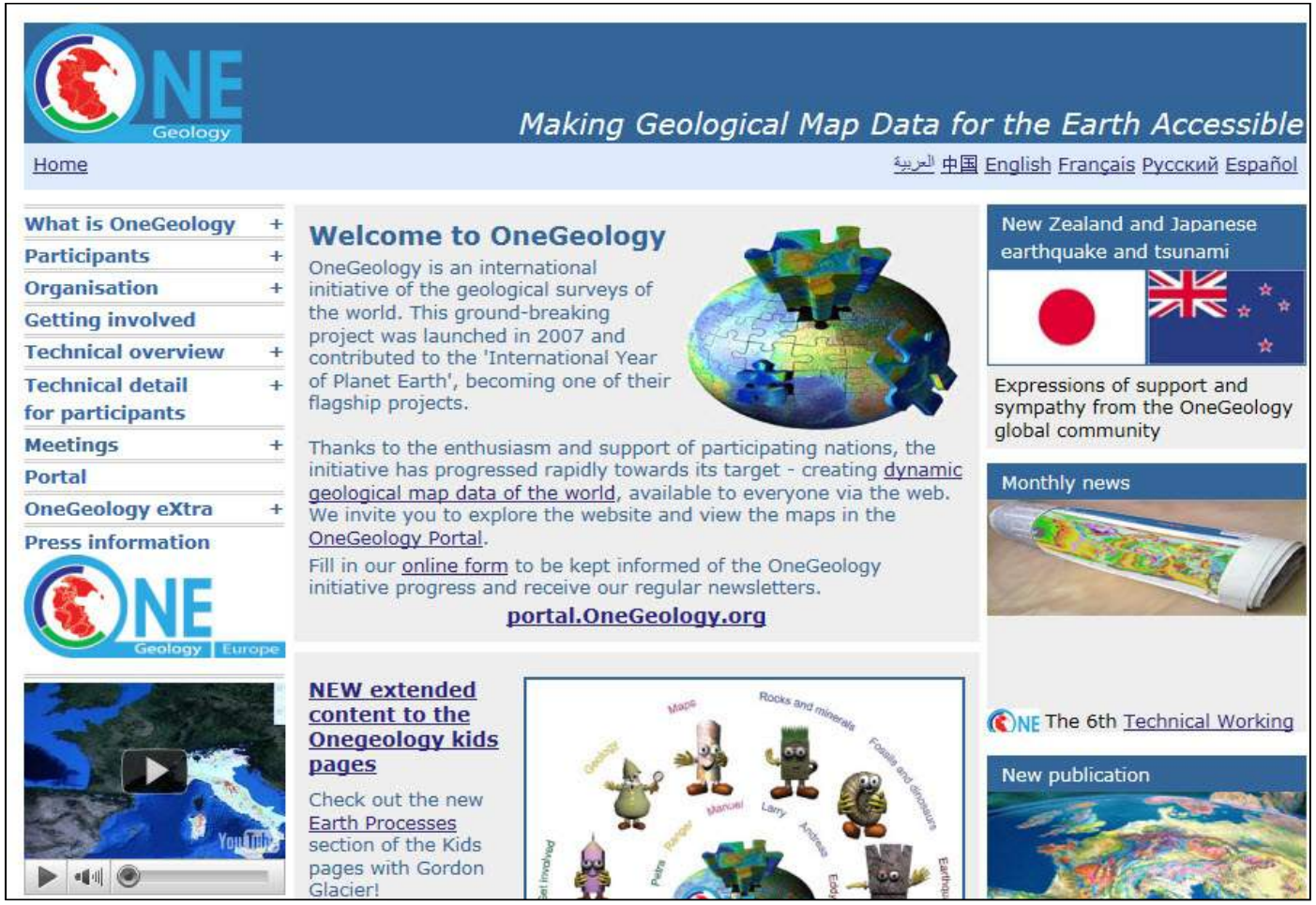
A presentation by
Clinton Smyth

cpsmyth@msn.com

* 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



OneGeology subscribes to GeoSciML and its Controlled Vocabularies

"Piloting EarthResourceML Applications in BC"

A 2011/2012 Geoscience BC Project

European Union

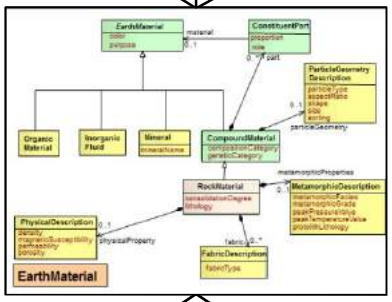
INSPIRE
Infrastructure for Spatial Information in Europe

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

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

Geoscience BC

United States



Approach: interoperability

- Overcome levels of heterogeneity

pragmatic	GW Practices (data usage)	
semantic	GW Ontology (data content)	
schema	GWML, WaterML (data structure)	Hydro/geology
syntax	GML (data language)	OGC
system	WFS, WMS,… (data systems)	

Natural Resources Canada / Ressources naturelles Canada
Canada

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**

Canada

Australia (Federal and State agencies)

All these are moving towards GeoSciML and its Controlled Vocabularies

Vocabulary Standardisation

A Critical Pillar to Computational Intelligence

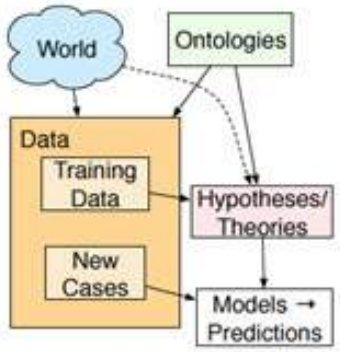
- Justification: Knowledge Interoperability
  - Standard Vocabulary is required for expressing human knowledge in computers
  - Standard Vocabulary is required for computer communication to other agents
- Resources Required
  - Vocabulary Repositories
  - Software Tools: Ontology and Taxonomy Editors; 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?

Geoscience BC

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/]
- 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)

Geoscience BC

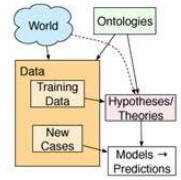
Resources Required to Enable Knowledge Interoperability

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

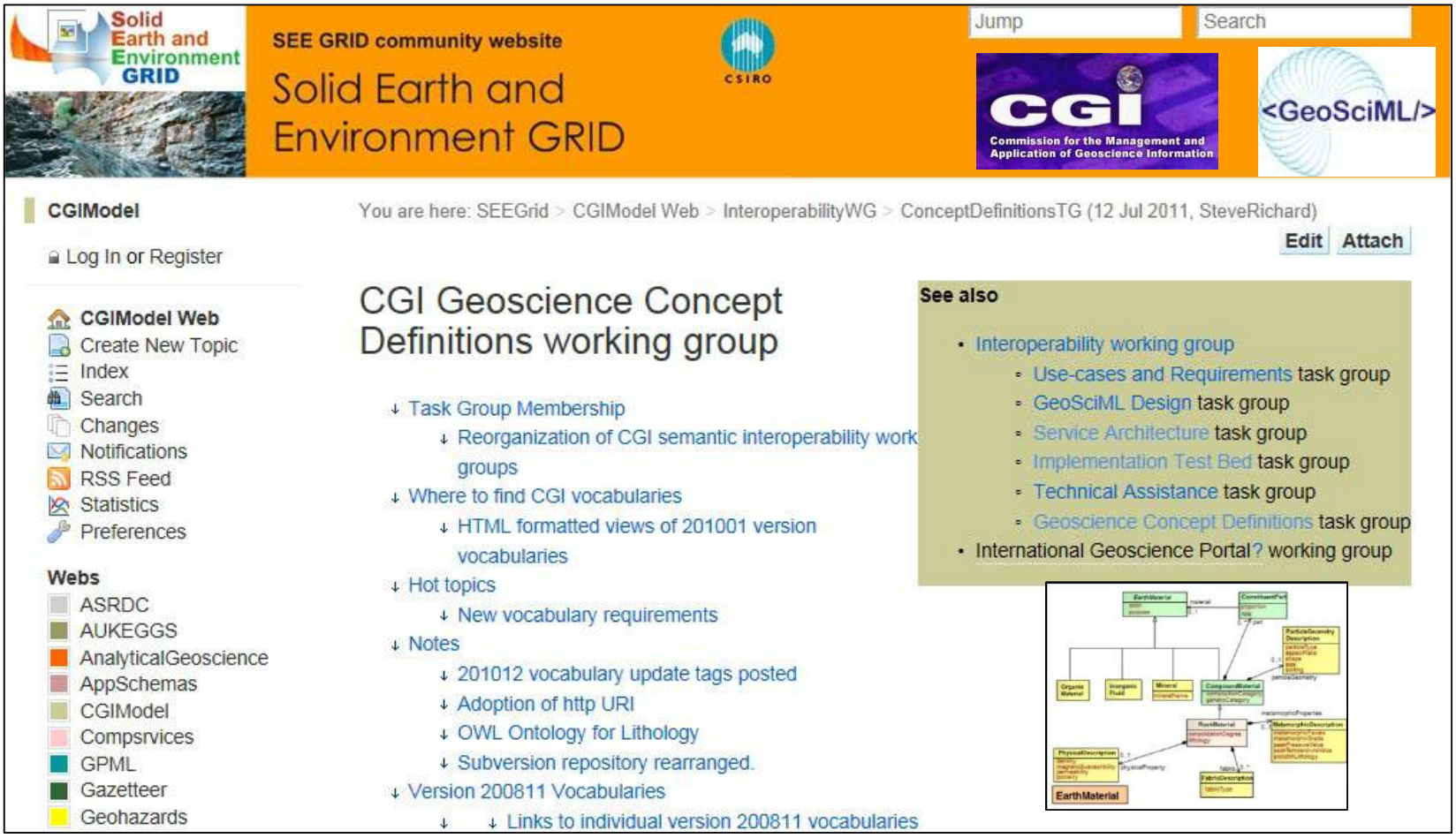
- Multi-Disciplinary Human Resources
  - Earth Science, Computer Science, Logic
    Probability, Semantics, Human-Computer Interfaces



- Promotional Applications
  - Without these, we cannot attract funding
    See: www.OneGeology.org

Global Vocabulary Repositories: GeoSciML

Enabling Global Interoperability while permitting Local Specialisation



http://onegeology.org/technical_progress/geosciml.html

Local Vocabulary Repositories

Maintaining Local Specialisation while enabling Global Interoperation

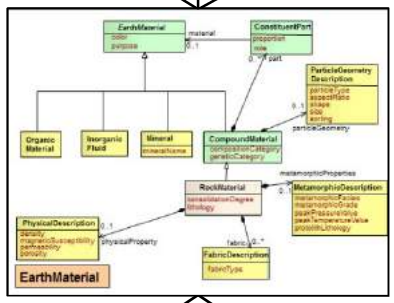
European Union

INSPIRE
Infrastructure for Spatial Information in Europe

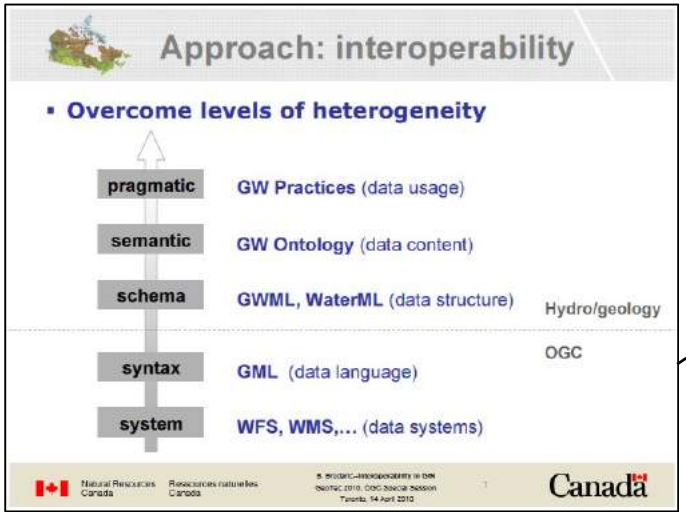
D2.8.III.21 Data Specification on Mineral Resources – Draft Guidelines

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

Geoscience BC



United States



Canada

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)

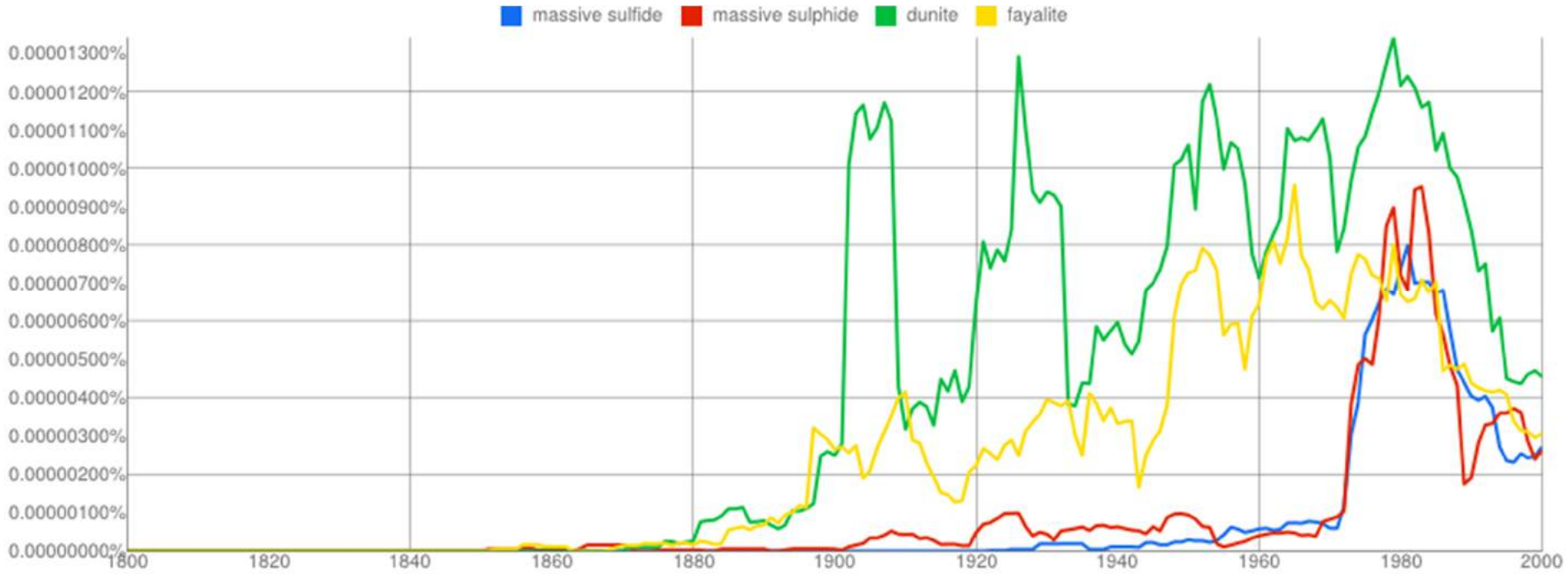
Vocabulary "Mediators" are the Key

http://en.wikipedia.org/wiki/Mediator_pattern

Local Vocabulary Repositories: Harvesting/Mediating Outcomes

Need to (constantly) revise the GeoSciML Taxonomy of Rocks

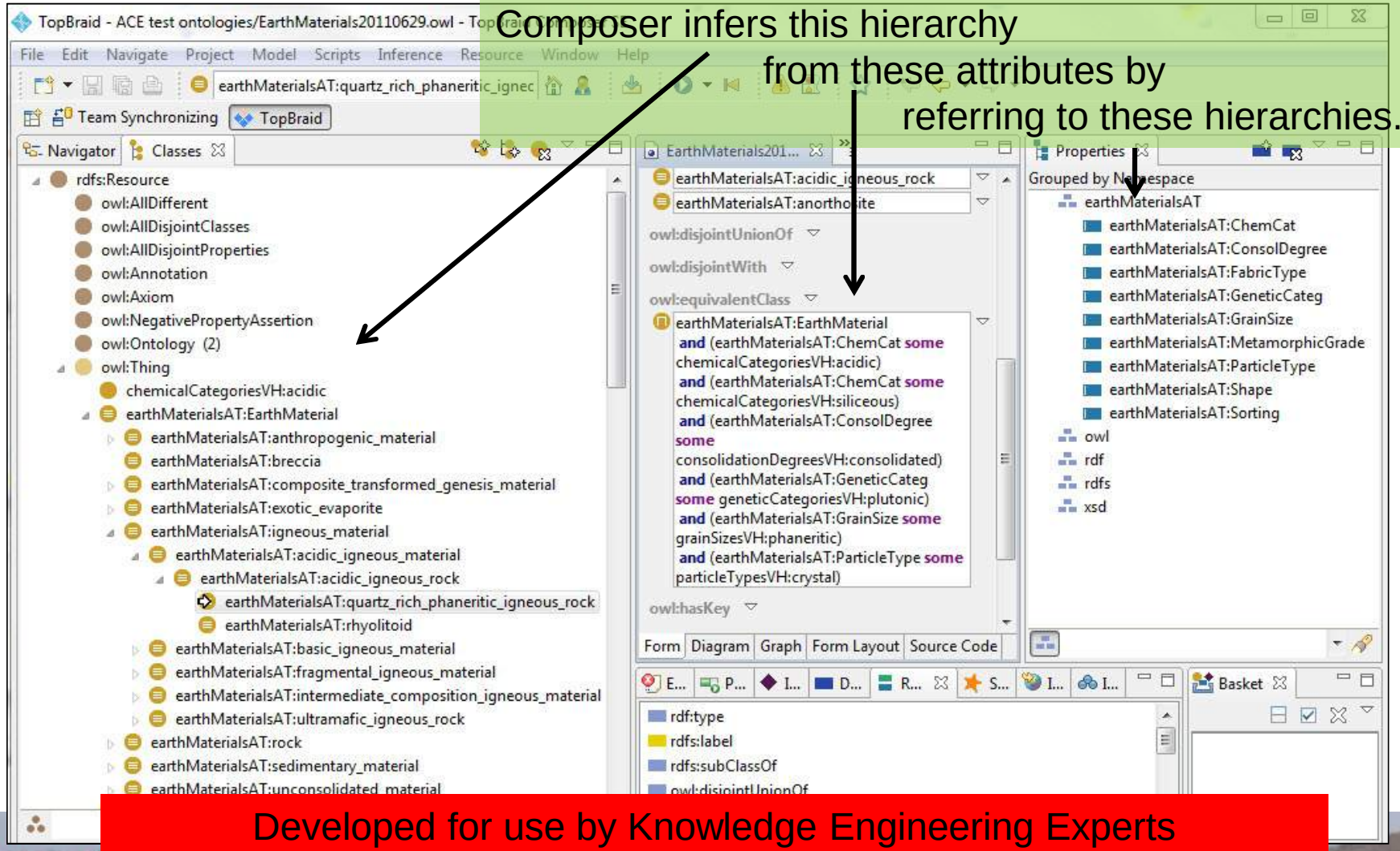
Graph below is a Google NGRAM (word-frequency over time) for Massive Sulfide, Massive Sulphide, Dunite, and Fayalite



"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

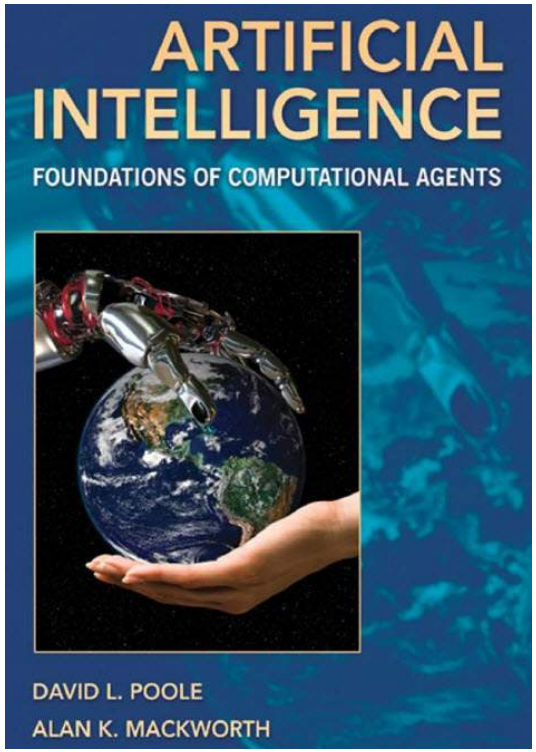
GeoSciML SimpleLithology201001 displayed in TopBraid Composer



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.

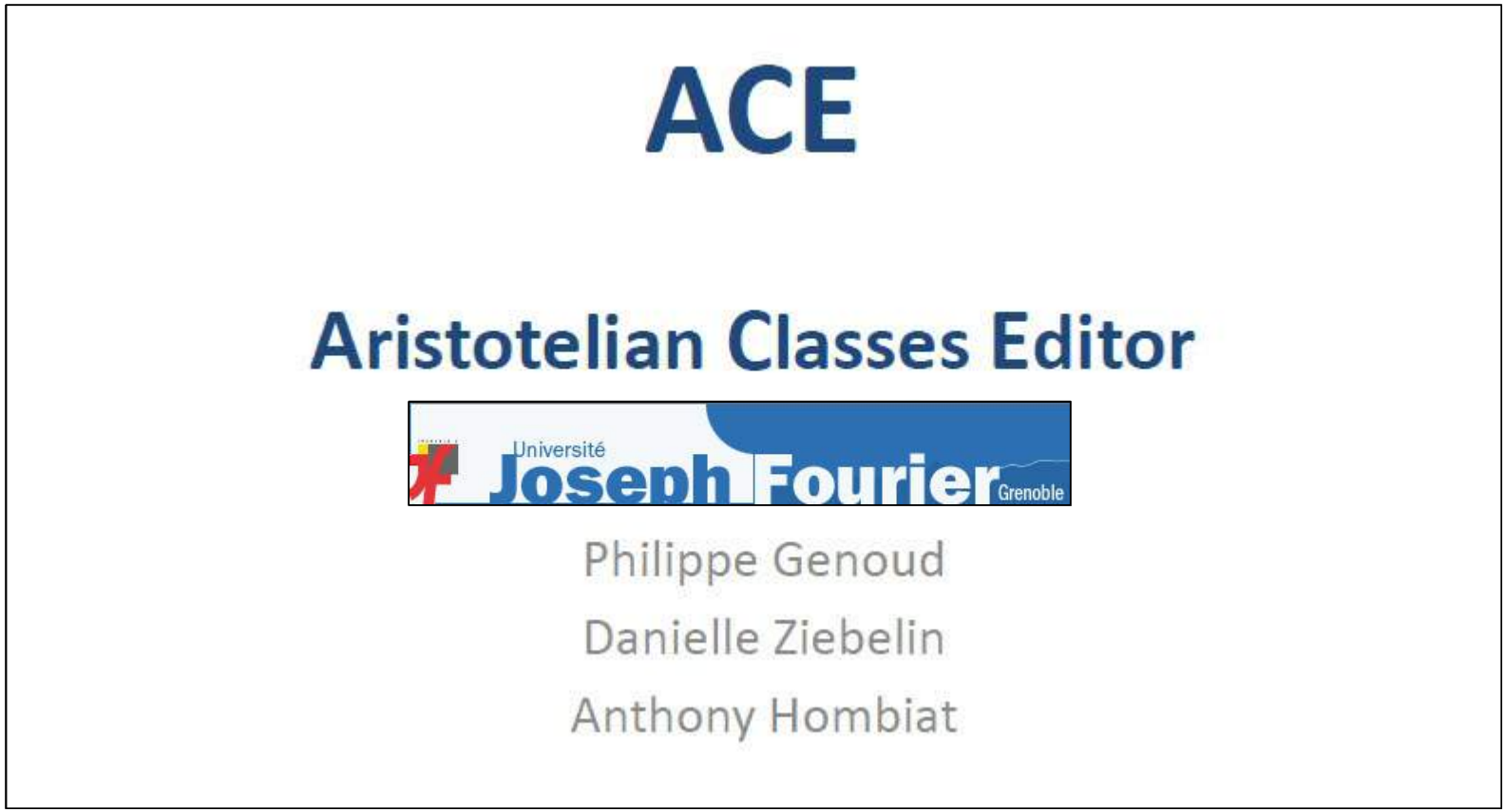
However, it is still straightforward to check whether one class is a subclass of another, to check the meaning of a class, and to determine the class that corresponds to a concept in your head.

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

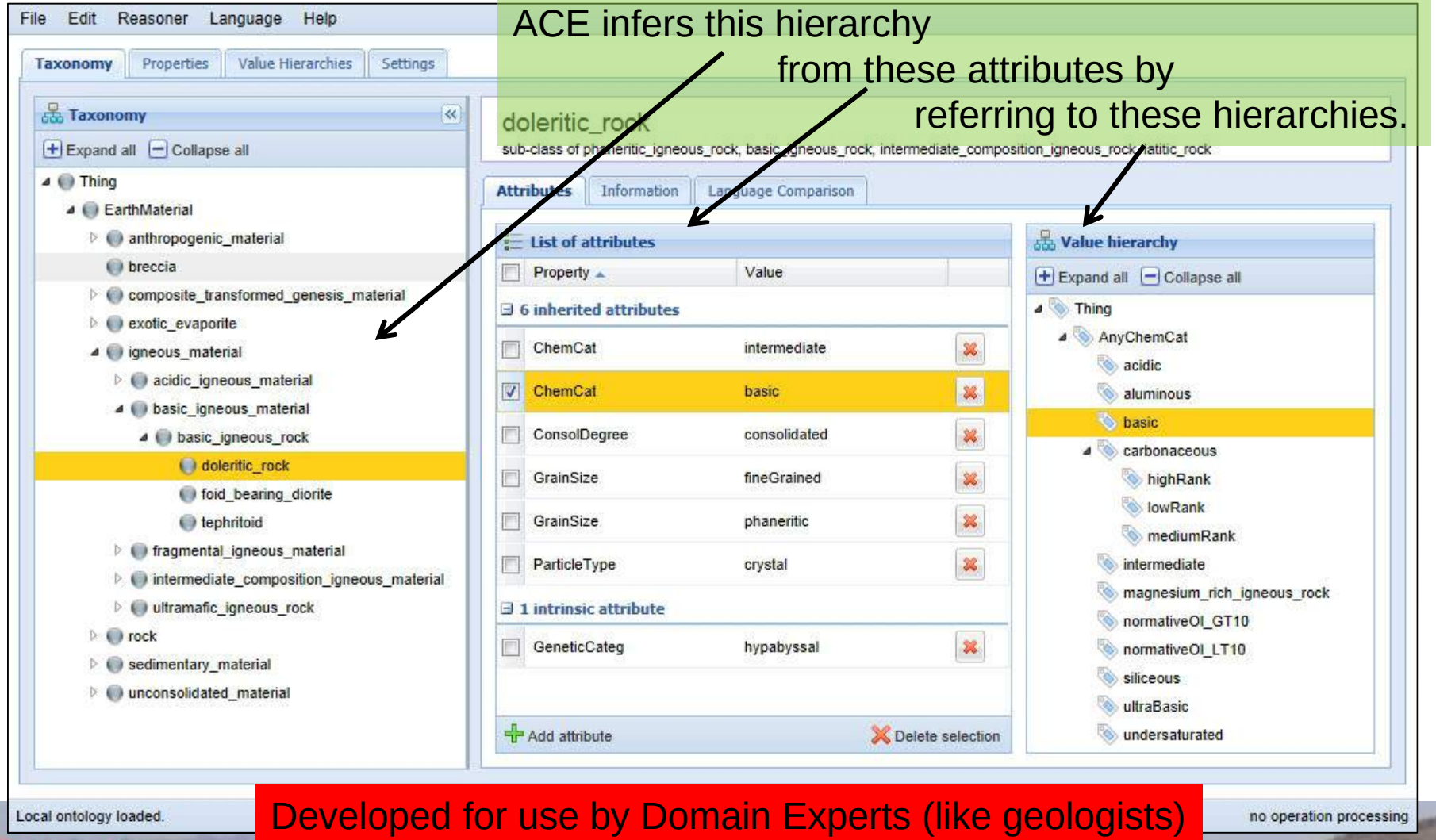
ACE: An Open-Source Taxonomy Editor
Developed by GOL and Université Joseph Fourier (Grenoble, France)



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

Vocabulary Software Tools: Taxonomy Editors

GeoSciML SimpleLithology201001 displayed in ACE

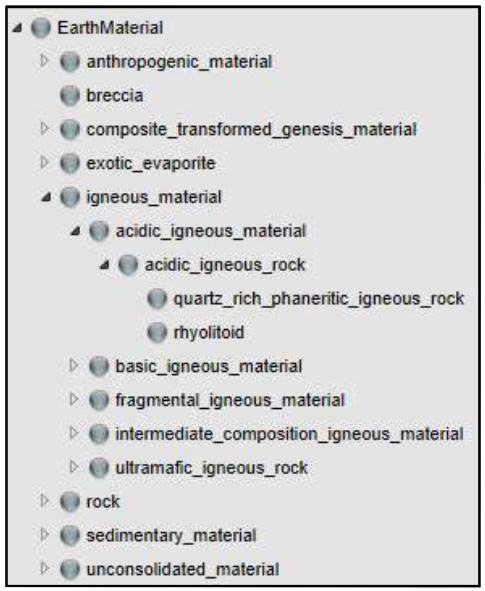


Developed for use by Domain Experts (like geologists)

GeoSciML SimpleLithology201001b: Excel Workflow

"Constituents" column was broken out to 16 independent "dimensions"

Inferred Taxonomy



	A	B	S	T	U
1	Rock Type	Constituents	ConsolDegree	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	carbonat
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	

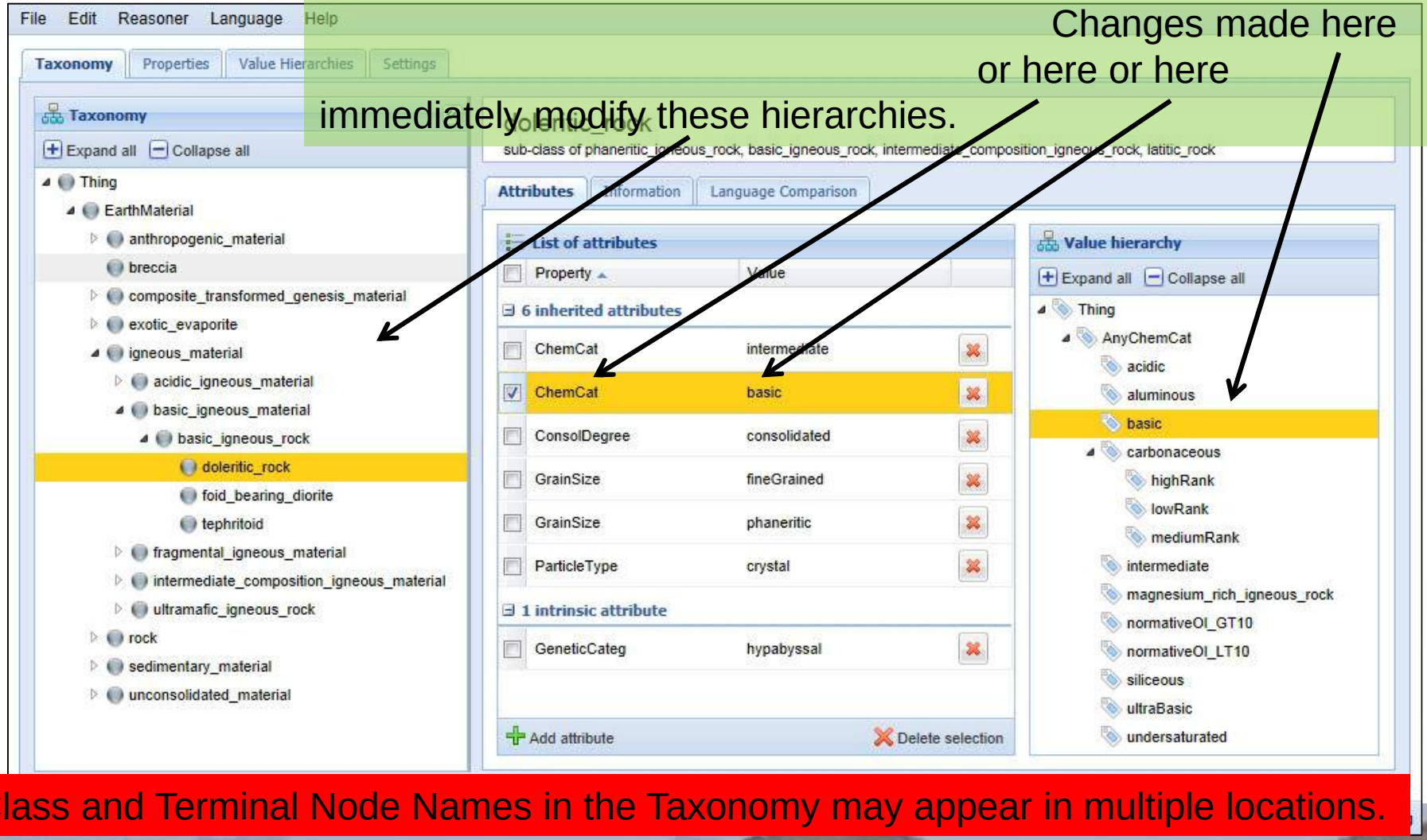
Pellet Reasoner

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

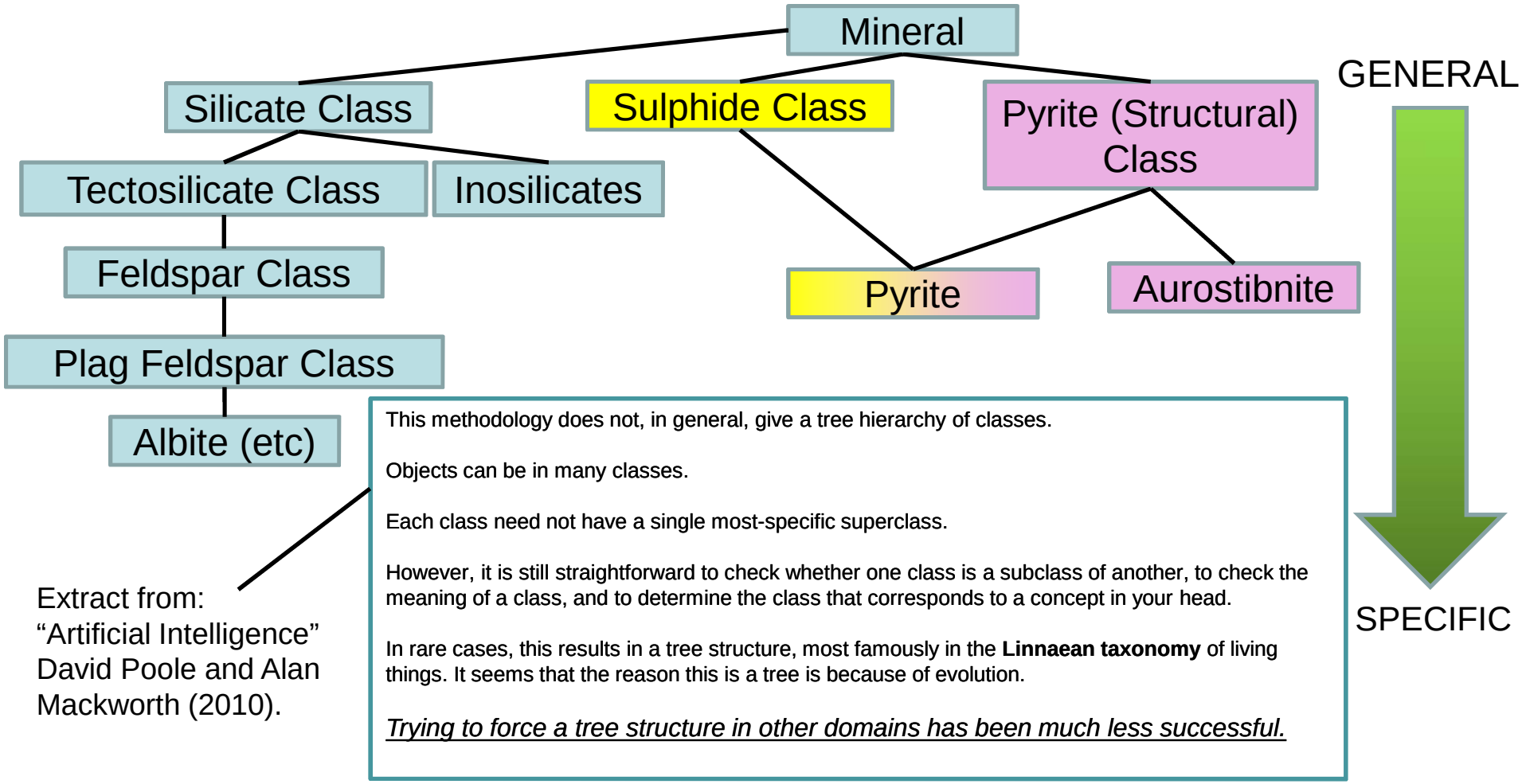
Changes made to Properties or their Values immediately update the Taxonomy



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



Extract from:
"Artificial Intelligence"
David Poole and Alan Mackworth (2010).

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.

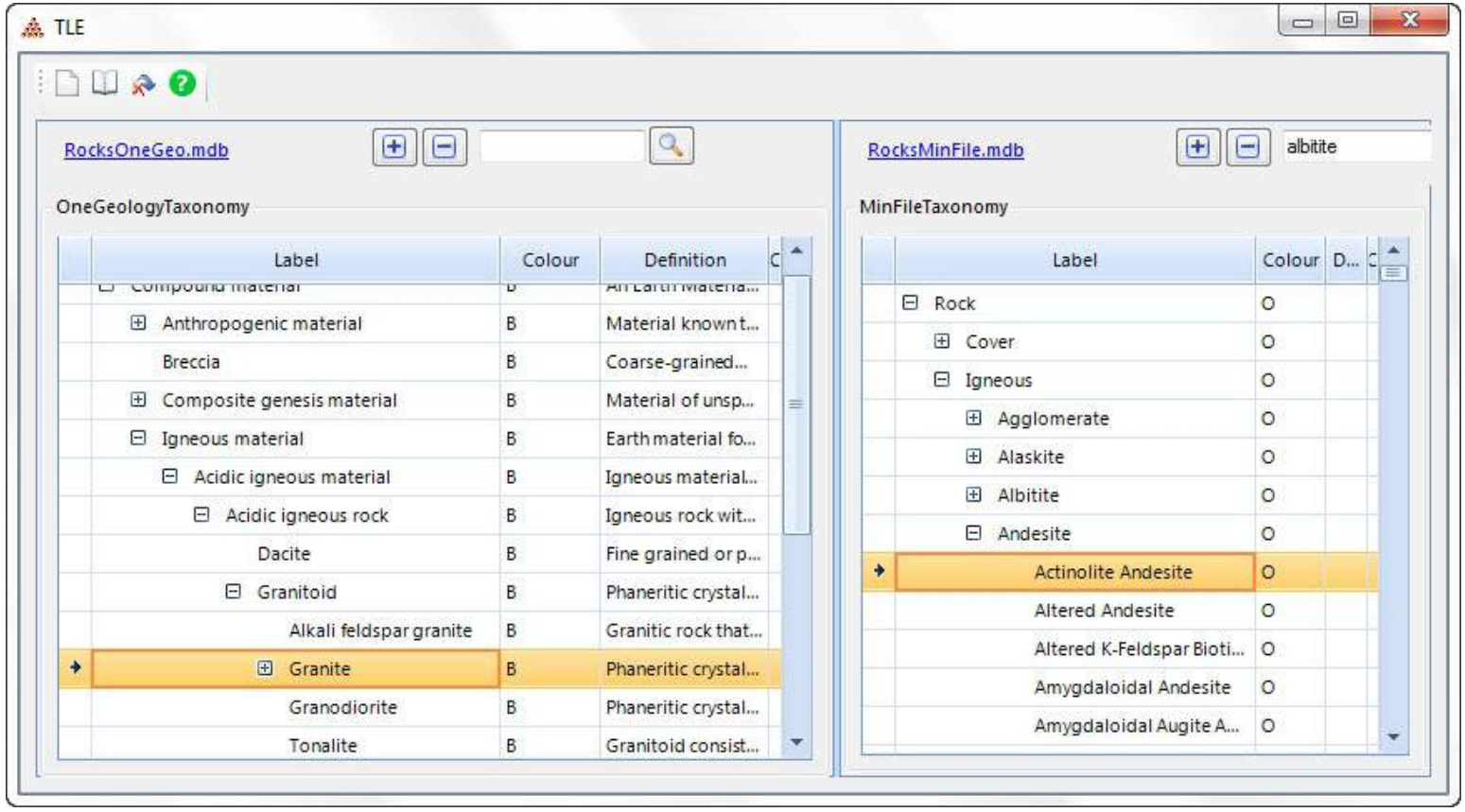
Trying to force a tree structure in other domains has been much less successful.

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

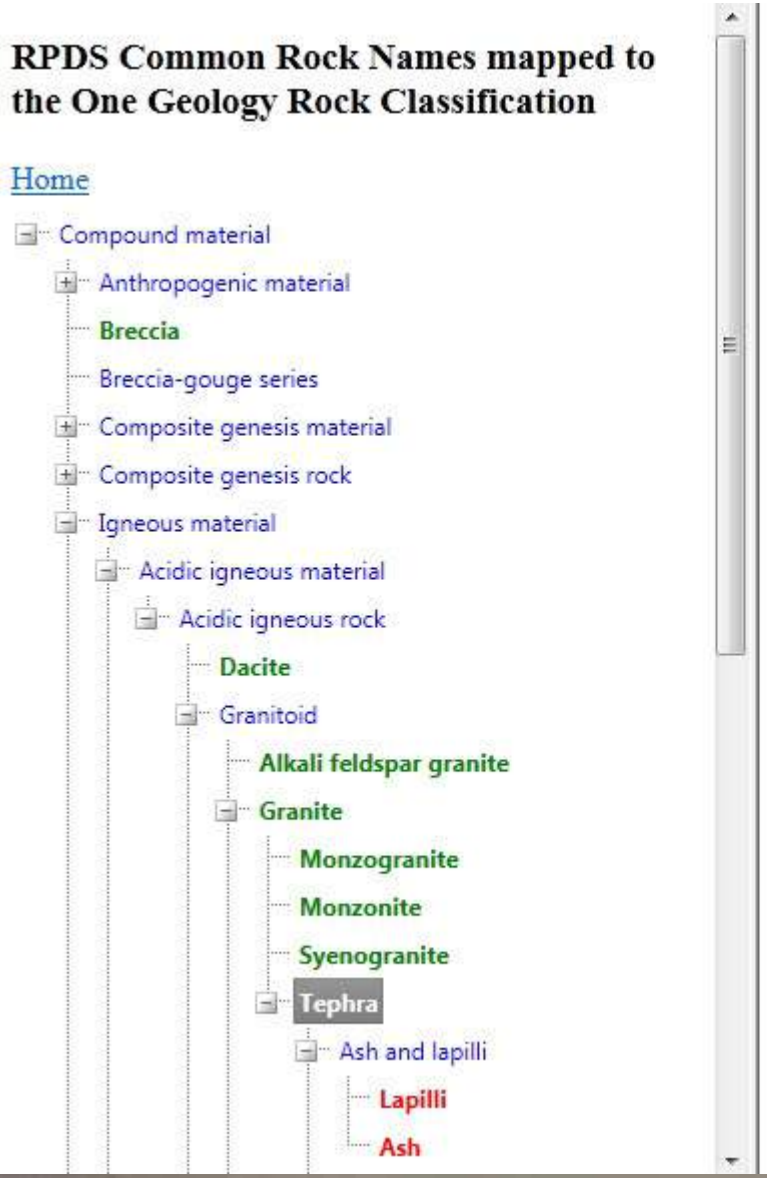
Easy-to-Use by Anyone Interested in Nomenclature

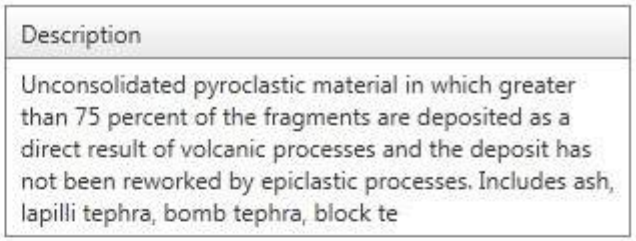


Available free from: www.golinfo.com/tle

Vocabulary Software Tools: Taxonomy Editors

TLE: TreeList Editor Output





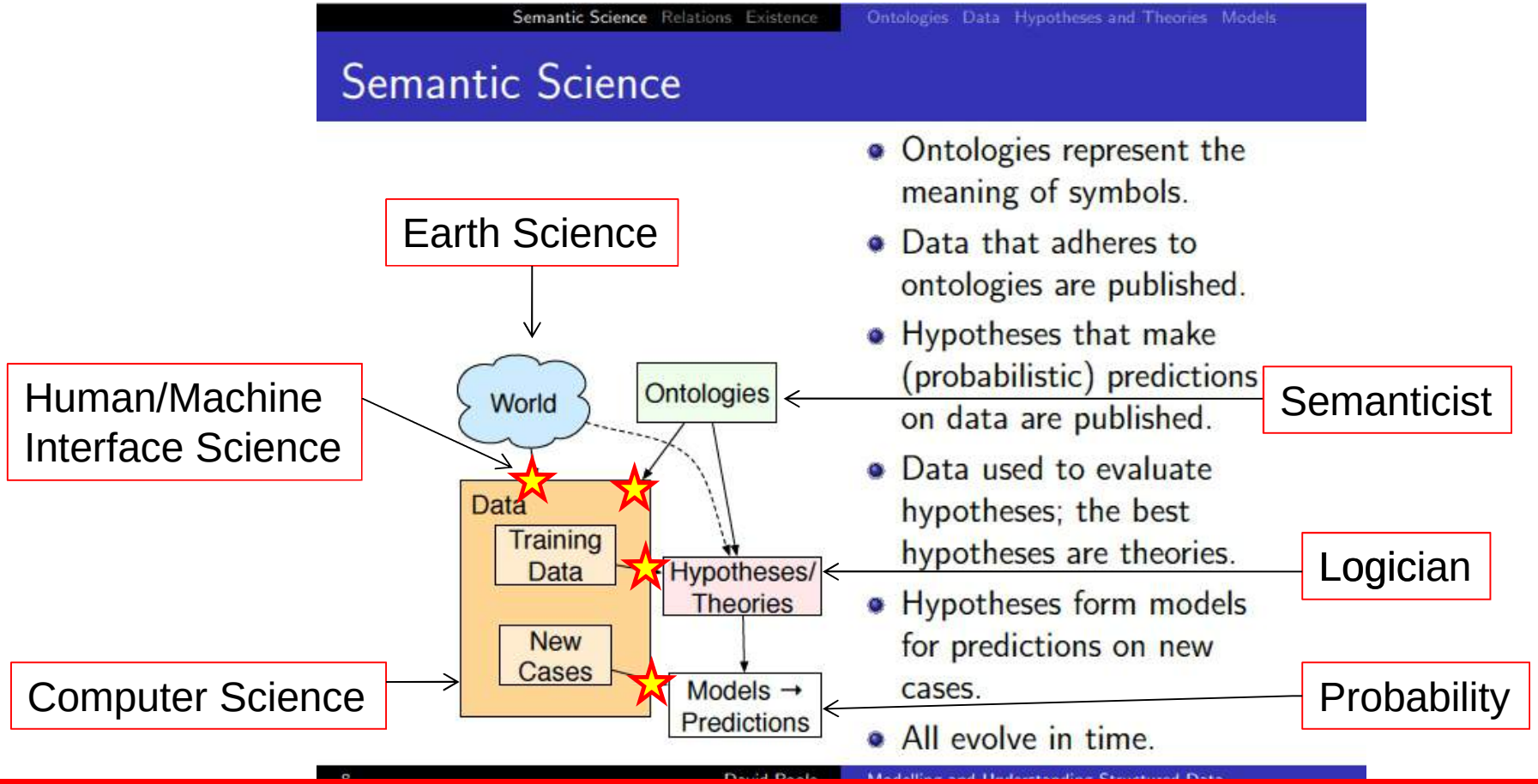
Facilitates Mapping between Taxonomies, and communicating the intricacies and importance of the subject

Geoscience BC

Vocabulary Development Human Resources

Must be Multi-Disciplinary

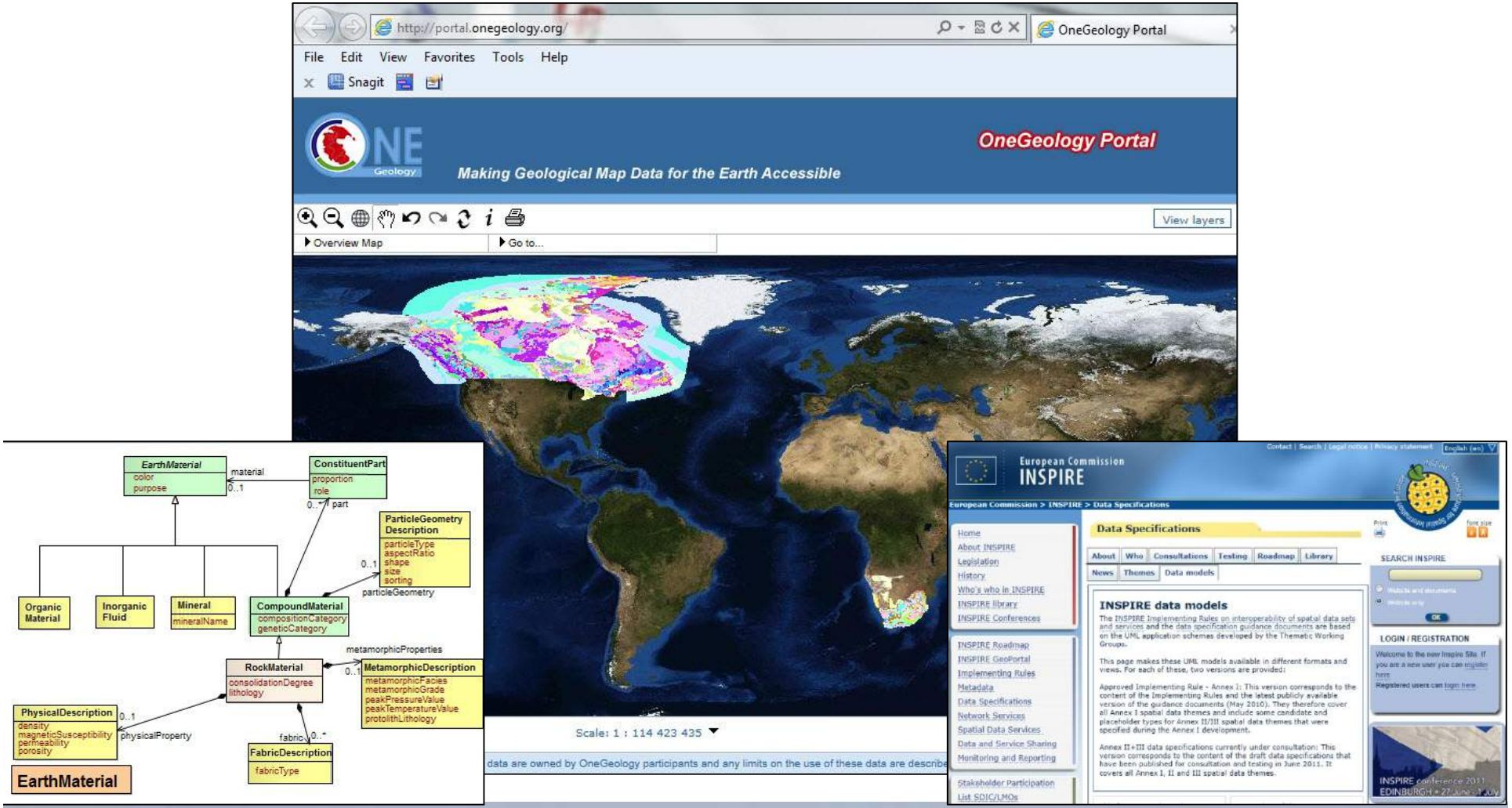
Absence of One Disciplinary Consideration leads to Flawed Vocabulary



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 21st Century – using them for better science



http://inspire.jrc.ec.europa.eu/

April 2012 Status: GeoSciML adoption in the US

OneGeology Workshop in Lakewood, CO

ONE Geology

Making Geological Map Data for the Earth Accessible

Home > Diary > USGIN/OneGeology Workshop

- What is OneGeology +
- Participants +
- Organisation +
- Getting involved
- Technical overview +
- Technical detail for participants +
- Meetings +
- Portal
- OneGeology eXtra +
- Press information

ONE Geology Europe

USGIN/OneGeology Workshop

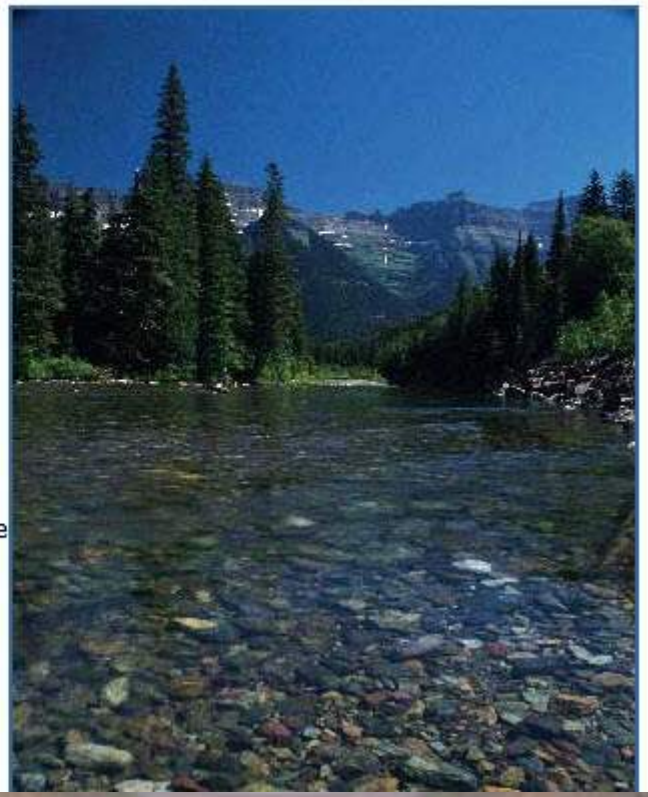
Dates: May 3 and 4, 2012 at USGS Ice Core Conference Room, Denver Federal Center, Lakewood, CO

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