Fostering the exchange of geoscience resources for knowledge exploration and discovery

Brandon Whitehead
The University of Auckland
Auckland, New Zealand
b.whitehead@auckland.ac.nz

Prof. Mark Gahegan
Centre for Digital Research
The University of Auckland
Auckland, New Zealand
m.gahegan@auckland.ac.nz

Melanie Evert
Chevron Energy Technology Company
Houston, USA
m.evert@chevron.com

Scott Hirs
Chevron Energy Technology Company
Houston, USA
djhirs@chevron.com

Boyan Broderic
Geological Survey of Canada
Ottawa, Canada
boyan.broderic@nrcan-rncan.gc.ca

ABSTRACT: Geoscience data capture is expensive. In order to extract maximum value, the data need to be consistently described, easily found, and then shared among those who need it. There has been recent momentum in the geoscience community to develop a common descriptive framework which facilitates data sharing. While storage and transfer standards are vital, they lack a descriptive element which standardizes the meaning of their contents. Metadata capture is appropriate for data stores, but often the terminology carries different meanings as domains become more specialized. For example, the term “migration” to a petroleum geoscientist refers to the movement of hydrocarbons in geologic time, yet to a seismologist it describes an imaging process. Furthermore, concepts associated with a term may change through time or as contextual factors in a discussion are modified. How, therefore, can the concepts evolved from geoscience resource terminology be defined and aligned to represent this multi-scaled orthogonal variability? Here we show how a community knowledge acquisition exercise was orchestrated to discuss fundamental concepts and their meanings as interpreted by leaders in basin characterization. The result of this exercise is a formal description of many of the features and processes associated with sedimentary basins, i.e. a basin ontology. This ontology allows the use of semantic connections between concepts as a backbone for search and discovery of research artifacts in large data stores.

1. An initial set of terms collected from source materials were used to spark discussion throughout the knowledge acquisition exercise.

2. Basin Characterization Experts review concepts and use images, documents and white boards to illustrate their concepts.

3. Knowledge acquisition team members dialogue with experts to disambiguate conceptual semantics.

4. Concepts and their relations are captured as a pseudo-formal representation in situ with the domain experts.

5. Concepts and relationships are codified using the Web Ontology Language (OWL) which allows for a semantic relationship to other concept spaces and description frameworks used in the domain, like the Semantic Web for Earth and Environmental Terminology (SWEET) and Geoscience Markup Language (GeoSciML) for example. The image illustrates a subset of the many connections between the Basin ontology created in this exercise (green nodes) and SWEET (tan nodes).

6. The ontology is then used to identify semantic connections in document collections. For example, these three documents share a relationship as two of the documents refer specifically to shale as a source rock, while the other refers to siliciclastic layers throughout the geologic time. This could be important during a search as shale is a type of siliciclastic rock.