



Geoscience terminology for data interchange

Stephen Richard

Interim Chair, CGI GTWG, Az. Geol. Survey, Tucson, Az, United States (steve.richard@azgs.az.gov)

Workgroups formed by the Commission for the Management and Application of Geoscience Information (CGI), a Commission of the International Union of Geological Sciences (IUGS) have been developing vocabulary resources to promote geoscience information exchange. The Multilingual Thesaurus Working Group (MLT) was formed in 2003 to continue work of the Multhes working group of the 1990s. The Concept Definition Task Group was formed by the CGI Interoperability Working Group in 2007 to develop concept vocabularies for populating GeoSciML interchange documents. The CGI council has determined that it will be more efficient and effective to merge the efforts of these groups and has formed a new Geoscience Terminology Working Group (GTWG, http://www.cgi-iugs.org/tech_collaboration/geoscience_terminology_working_group.html). Each GTWG member will be expected to shepherd one or more vocabularies. There are currently 31 vocabularies in the CGI portfolio, developed for GeoSciML interchange documents (e.g. see <http://resource.geosciml.org/201202/>). Vocabulary development in both groups has been conducted first by gathering candidate terms in Excel spreadsheets because these are easy for text editing and review. When the vocabulary is mature, it is migrated into SKOS, an RDF application for encoding concepts with identifiers, definitions, source information, standard thesaurus type relationships, and language-localized labels.

Currently there are 30 vocabularies still required for GeoSciML v3, and 38 proposed vocabularies for use with EarthResourceML (<https://www.seegrid.csiro.au/wiki/CGIModel/EarthResourceML>). In addition, a project to develop a lithogenetic map unit vocabulary to use for regional geologic map integration using OGC web map services is underway.

Considerable work remains to be done to integrate multilingual geoscience terms developed by the MLT Working Group with existing CGI vocabularies to provide multilingual support, and to make the thesaurus compiled by the MLT Working Group accessible online. The objective is to publish MTG/CGI concepts as RDF/XML using a standardized vocabulary service that would allow navigating asserted relationships between concepts in the CGI vocabularies and other vocabularies, e.g. a MTG/CGI concept is a skos:exactMatch to a GEMET concept, or a skos:closeMatch to a SWEET concept. Currently, the SISSvoc service developed by CSIRO Australia has been tested for deploying CGI vocabulary service and production deployment is planned for 2013. The existing vocabularies need to be established in permanent repository under stewardship of the GTWG, with documented policies for vocabulary management. The first formal meeting of the workgroup is planned for June, 2013.