

## Merging of Multidomain Vocabularies for Geoscience Applications

Bernd Ritschel, Germany/GFZ <http://www.gfz-potsdam.de>

The presentation shows the challenges and the benefits of merging SKOS modeled vocabularies related to geosciences and neighbor disciplines on the example of GCMD, GEMET and SPASE keyword classifications for annotation/retrieval applications and linked data purposes. The common usage of "standardized", domain related vocabularies for the description of data and information is a main premise for the realization of Semantic Web based applications and services.

NASA's Global Change Master Directory (GCMD) uses a well accepted vocabulary of hierarchical structured keywords for the description of more than 25,000 data sets for the earth and environmental sciences and related services. The General Multilingual Environmental Thesaurus (GEMET), which is published and maintained by the European Environment Information and Observation Network (EIONET), represents the standardized multi-language vocabulary (app. 6,000 terms, 27 languages) for the description and organization of environmental data of the European Environment Agency. The Space Physics Archive Search and Extract (SPASE) activity is a community based project for the data management of space and solar physics data which uses and maintains a standardized vocabulary for the description of data and related context information. For a sustainable combination and the shared use of the mentioned vocabularies a common data format is eligible. In the realm of the Semantic Web languages like the Resource Description Framework (RDF) or the Simple Knowledge Organization System (SKOS) are used. In our case, unfortunately only the GEMET vocabulary is provided in a SKOS modeled form. For the corporate usage, the keywords of GCMD and SPASE vocabularies had to be transformed into SKOS structures. After this activity a combination and merging of the different keyword vocabularies was possible. A big challenge for the combination of the vocabularies is the large number of terms in the GEMET thesaurus. So (half)-automated methods are necessary in order to find related terms between the different vocabularies as well as within the vocabularies itself. The challenges of the detection of equal terms or even parts of equal terms, the semantic proof of meaning and the choose of appropriate SKOS defined object properties and further qualification efforts represent the main part of this paper.