Geophysical Research Abstracts
Vol. 14, EGU2012-7374, 2012
EGU General Assembly 2012

From printed geological maps to web-based service oriented data products – strategies, foundations and problems.

M. Ebner, M. Schiegl, W. Stöckl, and H. Heger
Geological Survey of Austria, Vienna, Austria (marcus.ebner@geologie.ac.at)

The Geological Survey of Austria is legally obligated by the INSPIRE directive to provide data that fall under this directive (geology, mineral resources and natural risk zones) to the European commission in a semantically harmonized and technically interoperable way. Until recently the focus was entirely on the publication of high quality printed cartographic products. These have a complex (carto-)graphic data-model, which allows visualizing several thematic aspects, such as lithology, stratigraphy, tectonics, geologic age, mineral resources, mass movements, geomorphology etc. in a single planar map/product. Nonetheless these graphic data-models do not allow retrieving individual thematic aspects since these were coded in a complex portrayal scheme. Automatic information retrieval is thus impossible; and domain knowledge is necessary to interpret these "encrypted datasets". With INSPIRE becoming effective and a variety of conceptual models (e.g. GeoSciML), built around a semantic framework (i.e. controlled vocabularies), being available it is necessary to develop a strategy and workflow for semantic harmonization of such datasets.
In this contribution we demonstrate the development of a multistage workflow which will allow us to transform our printed maps to semantically enabled datasets and services and discuss some prerequisites, foundations and problems. In a first step in our workflow we analyzed our maps and developed controlled vocabularies that describe the thematic content of our data. We then developed a physical data-model which we use to attribute our spatial data with thematic information from our controlled vocabularies to form core thematic data sets. This physical data model is geared towards use on an organizational level but builds upon existing standards (INSPIRE, GeoSciML) to allow transformation to international standards. In a final step we will develop a standardized mapping scheme to publish INSPIRE conformant services from our core datasets. This two-step transformation is necessary since a direct mapping to international standards is not possible for traditional map-based data.
Controlled vocabularies provide the foundation of a semantic harmonization. For the encoding of the vocabularies we build upon the W3C standard SKOS (=Simple Knowledge Organisation System), a thesaurus specification for the semantic web, which is itself based on the Resource Description Framework (RDF) and RDF Schema and added some DublinCore and VoID for the metadata of our vocabularies and resources. For the development of these thesauri we use the commercial software PoolParty, which is a tool specially build to develop, manage and publish multilingual thesauri.
The corporate thesauri of the Austrian Geological Survey are exposed via a web-service that is conformant with the linked data principles. This web-service gives access to a (1) RDF/HTML representation of the resources via a simple, robust and thus persistent http URIs (2) a download of the complete vocabularies in RDF-format (3) a full-fledged SPARQL-Endpoint to query the thesaurus.
With the development of physical data-models (based on preexisting conceptual models) one must dismiss the classical schemes of map-based portrayal of data. E.g. for individual Geological units on traditional geological maps usually a single age range is given (e.g. formation age). But one might want to attribute several geologic ages (formation age, metamorphic age, cooling ages etc.) to individual units. Such issues have to be taken into account when developing robust physical data-models.
Based on our experience we are convinced that individual institutions need to develop their own controlled vocabularies and individual data-models that fit the specific needs on an organizational level. If externally developed vocabularies and data-models are introduced to established workflows newly generated and existing data may be diverging and it will be hard to achieve or maintain a common standard. We thus suggest that it is necessary for institutions to keep (or develop) to their organizational standards and vocabularies and map them to generally agreed international standards such as INSPIRE or GeoSciML in a fashion suggested by the linked data principles.