Design and Usage of Earth Science Ontology for Vertical Earth

Asanobu Kitamoto[1]; Yoshifumi Nogi[2]

[1] NII; [2] NIPR

http://agora.ex.nii.ac.jp/~kitamoto/

The cross-disciplinary integration of earth science data requires a reference model that facilitates mutual understanding among different conceptualizations in each domain. This paper discusses the design of ontology for earth science, and the usage of the ontology for our earth science database, Vertical Earth [1].

Earth science data has their commonalities that all data are about the Earth, but the contents are so diverse, and some are so long-term and huge that the cross-disciplinary integration of data is a daunting task. At the same time, however, cross-disciplinary information infrastructure is necessary for the understanding of the Earth system as a whole. We therefore focus on the vertical structure of the Earth system, namely X-spheres such as atmo-sphere, and challenge the problem of vertical integration of data divided into many spheres. This infrastructure may be useful also for the discovery of unexpected relationship between distant spheres.

For this purpose, we need a cross-disciplinary reference model for the Earth science data. Different conceptualizations in different domain make data integration a hard task. On the other hand, making a single uniform hierarchy for multiple domains is also a pessimistic task. So we initially focus on the integration of datasets by a model that allows the linking of multiple datasets. This model also helps users to find necessary datasets in unfamiliar domains by traversing the model through navigation or reasoning.

These types of reference models were already proposed. Firstly, we refer to GCMD (Global Change Master Directory) established by NASA. It is a meta-database with a comprehensive collection of earth science data from satellites and in-situ observations, and this already provides metadata of more than 20,000 datasets. This is practically useful as an integrated database of earth science data, but vocabularies and hierarchies are not well organized. This problem is tackled by another project SWEET (Semantic Web for Earth and Environmental Terminology) also at NASA. This is an earth science ontology proposed in 2006, and built on OWL (Web Ontology Language). SWEET Ontology seems to be the most comprehensive ontology on earth science data, so we refer to this ontology to build our ontology that focuses on the vertical structure of the earth. The tool for designing ontology is Protege, and the language is OWL, which is the same language as SWEET.

This ontology will be used for the vertical integration of earth science data at our site, Vertical Earth. In the first step, ontology is used for metadata level, which means the integration of datasets by the characterization of the whole datasets. This level is almost the same as GCMD, but the integration will be more systematic by using the ontology. The second step will be integration at the data level, which means the overlaying of data through unit conversions, data corrections, data projections, and so on. This involves diverse processing methods and hence is hard to provide a general solution, but we start from practical problems by implementing Web services for multiple earth science data.

Future work includes the application of ontology for Vertical Earth, such as GPV atmosphere data or Antarctic geographic data, and continues interoperability tests for earth science data.

Acknowledgment: This work is supported by research grants from Transdisciplinary Research Integration Center, Research Organization of Information and Systems.

[1] Vertical Earth, http://earth.nii.ac.jp/