

Vertical Earthにおける地球科学オントロジーの設計と活用

Design and Usage of Earth Science Ontology for Vertical Earth

北本 朝展 [1]; 野木 義史 [2]

Asanobu Kitamoto[1]; Yoshifumi Nogi[2]

[1] 国情研; [2] 極地研

[1] NII; [2] NIPR

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

地球科学データを分野横断的に統合するためには、分野ごとに異なる概念体系の間で相互理解を促進するための参照モデルが必要になる。そこで本論文では、地球科学データを対象としたオントロジー (ontology) を設計し、これを我々が構築する地球科学データベース Vertical Earth [1] で活用する。

地球科学データは、地球に関するデータであるという共通項は持つものの、その内容は分野ごとに多様であり、しかも長期・巨大なデータも多いことから、分野横断的なデータ統合は困難な状況にある。しかし地球というシステムを総合的に理解するためには、様々な分野のデータを統合した情報基盤を整えなければならない。そこで我々は地球科学データの特徴として地球システムの鉛直構造 (圏) に着目し、圏ごとに分断されたデータを統合するという「鉛直統合」の問題に着目する。これは、圏ごとに生成された複数データから、圏をまたいだ「遠隔層間の意外な関係」をデータに基づいて発見するための情報基盤になりうるものである。

そこで必要となるのが、分野横断的な地球科学データの参照モデルである。地球科学の分野ごとに概念体系が異なったままでは、分野横断的にデータ統合を進めることは困難である。しかし、学問体系そのものを統合するのはあまりに困難な作業となるため、まずは地球科学データ全体を俯瞰的に眺め、データの間を明示的に定義できるモデルについて考える。個々のデータセットがそのモデルにリンクされるようになれば、データセットの関係をモデル上で推論することもできるため、自分が不案内な分野においてもユーザが必要なデータセットにたどりつける可能性が高まることになる。

こうした共通的なモデルはすでにいくつか構築されている。まず NASA による GCMD (Global Change Master Directory) は、衛星観測や現場観測において得られる地球科学データを包括的に収集したメタデータベースであり、すでに2万を超えるデータセットに関するメタデータを収集している。これは実用的な地球科学データの統合データベースとして価値あるものであるが、メタデータに利用する語彙体系が整理されておらず、階層構造に統一的な基準が見られないという問題がある。そうした問題を解決するために提案されたのが NASA の SWEET (Semantic Web for Earth and Environmental Terminology) である。これは2006年に提案された地球科学オントロジーであり、OWL (Web Ontology Language) を用いて概念間の関係なども明確に定義されていることから、関係に基づく推論も可能となっている。この SWEET オントロジーは地球科学に関する現時点で最も包括的なオントロジーであると考えられることから、本論文でもこのオントロジーを利用しつつ、特に地球の鉛直構造に焦点を合わせたオントロジーを設計する。オントロジーの設計ツールには Protege を利用し、オントロジー言語には SWEET と同様に OWL を利用する。

こうして構築したオントロジーは、著者らの Vertical Earth におけるデータ統合に利用する。第一段階ではメタデータのレベル、すなわちデータセットの全体的な特徴に基づく統合を目標とする。これは GCMD と類似したレベルであるが、語の間の関係をオントロジーで定義することにより、ユーザはあるデータセットと別のデータセットにどのような関係があるのかを理解しながら探すことができるようになる。第二段階ではデータレベルの統合、すなわち個々のデータの重ね合わせなどによる統合を目標とする。ここでは単位変換やデータ補正、地図投影変換など、基本的な部分に限っても多様な処理が介在するため一般的な解決は困難ではあるもの、ウェブサービスとして実現可能な範囲から実践的に解決していく。

今後は本研究で構築した地球科学オントロジーを、著者らが Vertical Earth にて構築を進めている大気 GPV データや南極地理データなどに適用し、地球科学データの相互運用に向けて実験を進めていく計画である。

謝辞：本研究は、情報・システム研究機構、新領域融合研究センタープロジェクトとして支援を受けている。

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

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/>