

Modeling with the Semantic Web in the Geosciences

Femke Reitsma

Institute of Geography, School of Geosciences, University of Edinburgh
femke.reitsma@ed.ac.uk

Jochen Albrecht

Department of Geography, Hunter College, CUNY
jochen@hunter.cuny.edu

Earth system science deals with complex systems that pose many significant representation challenges. As depicted in the classic Bretherton diagram of biospheric cycles (Figure 1), modeling the earth system involves numerous interacting components, each of which can be further dissected into sub-components that are studied by specialists in a wide range of disciplines.

From this description, problems of both model interoperability and the model simulator interoperability already become evident. Given the complexity of the task and the number of research groups and individuals involved, there exist a wide diversity of modeling approaches, such as models based on differential equations or stochastic methods, that make not only the interoperation of model specifications difficult but the intercomparison of the structure and results of similar models as well, as is evident in the work undertaken by The Global Analysis, Integration and Modeling Task Force (GAIM) (for an example see¹). Similarly, in terms of simulator interoperability there is also a wide range programming languages and software in which models are developed, making it difficult to couple a FORTRAN model of thermohaline circulation with an ice sheet model in C++.

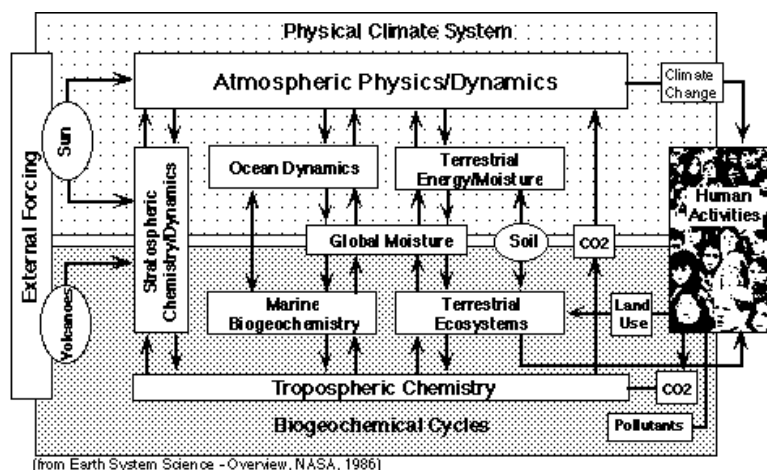


Figure 1. The Bretherton diagram

Source: Earth Systems Science Overview: a program for global change, NASA 1986

Furthermore, issues surrounding spatial data compound these issues. Spatial data form one of the primary inputs for models, and as with all other types of data, its volume continues to grow at an explosive rate². Yet there is a worldwide trend of declining use, management, and content of national clearinghouses for spatial data, due to a dissatisfaction of the spatial data community with the functional capability of these portals providing spatial data³.

Furthermore, much of the knowledge about the physical systems that are modeled lies, from a computing perspective, dormant in scientific papers, modeling code, and in the heads of scientists. Ontologies as knowledge repositories have been developed to support the primary goal of sharing

knowledge in a manner that aids understanding⁴. However the development of ontologies for geoscience disciplines has been limited to keyword lists for classification, such as the Global Change Master Directory's (GCMD) earth science keywords⁵, or ontologies that are essentially class hierarchies with some limited expression of properties, such as NASA's Semantic Web for Earth and Environmental Terminology (SWEET) ontologies. The potential of ontologies and the semantic web has yet to be tapped for scientific modeling and simulation.

These problems largely derive from a common lack of explicit semantics in representing models⁶, spatial data, and scientific knowledge in general.

Process Ontologies

For modeling earth system processes, ontologies are needed in order to develop conceptually sound models, effectively communicate these models, enhance interoperability between models developed in different domains, and provide the opportunity for reuse and sharing of model components. To accomplish these goals, these processes need to be expressed not only in terms of their types and properties, but in terms of their behavior, spatial and temporal characteristics, relationships to other processes, data requirements for implementation, and spatial data models for visualization and storage of results. A collection of such process descriptions could then form the basis for a library of processes to be utilized by a simulation framework.

Modeling Process Behavior

What is needed is an extended notion of ontology that can express rules defining the thresholds of process change, and operations expressing the behavior of the process. An example of a likely rule is: if variable *x* has a wind speed greater than 65 knots, and *x* is located in a place (represented by variable *y*) called the "Western Pacific", then *x* is a "Typhoon". The consequent method might then call on a set of typhoon processes to be initiated, which in turn interact with other processes, such as coastal erosion. Thus we need specifications that not only describe *what* the processes of the model are, but *how* those processes operate. Ultimately this will support the interaction of model components at the level of process definitions, that is, interoperation at the level of the model components rather than interoperation as input and output to the model, which is the current dominant approach.

DARPA is currently developing an extension to OWL called SWRL (Semantic Web Rule Language), which allows for the expression of some aspects of rules and behavior of processes. SWRL injects into OWL parts of RuleML (Rule Markup Language), thereby extending the set of OWL axioms to include Horn-like rules. Thus far, SWRL only includes a restricted part of this abstract rule type, namely the derivation rules, which assert a conclusion when certain conditions hold. To represent behavior, SWRL can change the values of properties of classes or can call an external "oracle" with *BuiltIns*. SWRL BuiltIns have been developed for future extensions of the language, and are essentially calls to an external method or program that returns information required to evaluate the SWRL statement. BuiltIns can be used to incorporate programmed behavior by calling a program or implementation of some process behavior.

Ontology-based Simulation

With an ontology that expresses all of the relevant dynamic features of the processes to be modeled, such as watershed runoff, ocean heat transport, or atmospheric circulation, we can compose these processes into a simulation framework. With the research directed at converting semantic web languages to running code^{7,8} this looks increasingly possible. Ontology-based simulators should allow for reasoning with process descriptions enabling us to determine whether the other model components are available to make a complete model. Ideally, this will lead in the future to simulation

platforms that assist in determining whether the logic of the process description is correct according to scientific knowledge bases.

The Semantic Web will further facilitate collaboration of researchers and the automated discovery and use of model components. Modeling the earth as a system requires an enormous breadth of knowledge of physical processes, a knowledge base that no individual scientist holds. Even at the level of identifying what processes are important to model in certain systems requires the collation of knowledge of a large array of researchers. By expressing the semantics of what aspect of an environmental system a process description models and how it interacts with other processes, the future web may be peppered with process definitions that scientists have logged on their websites for automated modeling agents to discover and use.

Utilizing the Semantics of Spatial Data

Spatial data forms a key role in modeling the earth system as the input to models as well as a measure against which results are validated. However, the metadata for spatial data is in general in a poor state due to the lack of interest in creating it at the time of data collection or data modification. The lack of metadata may limit discovery of spatial data for modeling, but even with its full expression, it limits the search for spatial data to predefined keywords and is further hampered by the approach to presenting that data online.

Gazetteers, such as the World Wide Gazetteer or The Alexandria Digital Library (ADL) Project's Gazetteer service, are the most common approach to geographic information retrieval. They allow us to search for geographic features such as cities, deserts, and jungles, or the location of other information based on the spatial location or attributes of those geographic features. Implementations of gazetteers or gazetteer centered search engines are based upon a query submitted to the system either interactively or through an API for large-scale data retrieval. This approach can often limit accessibility to interacting with the geographic data in predefined ways, which are specified according to the capabilities of the software and the relationships defined within the database. This is particularly evident with the proliferation of spatial data providers, which all have unique methods of spatial data extraction (for both human and computer agent) resulting in a significant amount of time spent learning the structure of the data provider.

Such portals typically adhere to syntactic metadata standards, such as the U.S. federally mandated Federal Geographic Data Committee's (FGDC) Content Standard on Digital Geospatial Metadata (CSDGM) and the ISO 19115 Geographic Information Metadata standard and increasingly syntactic spatial data standards, most notably GML 3.0 (Geographic Markup Language). More recently the semantics of these standards have been expressed in a set of ontologies produced at Drexel University. Yet there remain no accepted semantic standards for expressing the formal semantics of spatial data or the metadata describing that data⁹ (Bernard *et al.*, 2005).

Rather than a geoportal being a website where geographic content can just be discovered¹⁰, the solution is ontology-based discovery and retrieval of geographic information¹¹. However, beyond the expression of the semantics of metadata for enhancing data discovery and use of spatial data, the semantics of the spatial data itself, in the form of geographic features such as coastlines and buildings need to be expressed for automated discovery and use. For example, if a user was interested in finding spatial data that had a climate station within a specific watershed, the information to answer this query is expressed in the attribute tables of the data and as a relationship between two different data sets, which is not available in the metadata or the functionality of spatial data portals. Providing access to this information and its semantics will also enhance the use of model results for automated analysis and use.

Towards Modeling the Earth on the Semantic Grid

Beyond meeting the specific needs of modeling the earth's systems and providing semantic solutions for interoperating among models and simulators, understanding of that system needs to be expressed and stored in a knowledge base. This will soon provide the bottleneck for utilizing Semantic Web technology for scientific research. This has been partly recognized by a number of recent large scale initiatives, such as the Science Environment for Ecological Knowledge (SEEK) project and the GEOsciences Network (GEON) project, which intend to create ontologies for their specific domains and utilize them on new large scale platforms that will provide access to services for modeling and analysis and spatial data warehouses. These projects aim at developing a *cyberinfrastructure*, a Semantic Grid for science.

Expressing and utilizing semantics in modeling the earth will enhance our ability to do science and could lead to new insights to the environmental systems that are studied and a greater understanding of existing methods that are used to represent those systems. The Semantic Web provides the platform for developing the solutions described above and provides the opportunity to utilize the models and modeled results in new and interesting ways. For example, the challenges of natural disasters, such as the recent tsunami in Asia, highlights the potential utility of integrating models with spatial decision support systems (SDSS) for timely response to extreme events.

Related Links

Alexandria Digital Library (ADL): <http://www.alexandria.ucsb.edu/gazetteer/>
FGDC standard: <http://fgdc.er.usgs.gov/metadata/constan.html>
GAIM: http://gaim.unh.edu/about_gaim.html
GCMD: <http://gcmd.gsfc.nasa.gov>
GEON project: <http://www.geongrid.org>
Drexel University's geospatial ontologies: <http://loki.cae.drexel.edu/~wbs/ontology/list.htm>
GML: <http://www.opengis.org/docs/02-023r4.pdf>
ISO19115 <http://twww.isotc211.org/scope.htm#19115>
RuleML: <http://www.ruleml.org/>
SEEK: <http://seek.ecoinformatics.org/>
SWEET Ontologies: <http://sweet.jpl.nasa.gov/ontology/>
SWRL: <http://www.daml.org/rules/proposal/>
World Wide Gazetteer: <http://www.gazetteer.com/>

References

-
- ¹ Dutay, J. C., J. L. Bullister, S. C. Doney, et al., 2002. Evaluation of Ocean Model Evaluation of ocean model ventilation with CFC-11: comparison of 13 global ocean models Ventilation with CFC-11: Comparison of 13 Global Ocean Models. *Ocean Modelling*, 4: 89-120.
- ² Lu, C. T., Y. Kou, H. Wang, et al., 2003: Two Web-based Spatial Data Visualization and Mining Systems: Mapcube & Mapview. *International Workshop on Next Generation Geospatial Information*, Cambridge (Boston), Massachusetts.
- ³ Crompvoets, J., A. Bregt, A. Rajabifard, and I. Williamson, 2004. Assessing the Worldwide Developments of National Spatial Data Clearinghouses. *International Journal of Geographical Information Science*, 18: 665-689

⁴ Gruber, T. R., 1993. A Translation Approach to Portable Ontologies. *Knowledge Acquisition*, 5: 199-220.

⁵ Olsen, L. M., G. Major, S. Leicester, K. Shein, J. Scialdone, H. Weir, S. Ritz, C. Solomon, M. Holland, R. Bilodeau, T. Northcutt, and R. Vogel: NASA/Global Change Master Directory (GCMD) Earth Science Keywords. Version 4.2.2. Available online from http://gcmd.gsfc.nasa.gov/Resources/valids/keyword_list.html

⁶ Mackay, D. S., 1999. Semantic Integration of Environmental Models for Application to Global Information Systems and Decision-Making. *ACM-SIGMOD*, 28: 13-19.

⁷ Kalyanpur, A., D. Partor, S. Battle, and J. Padget, 2004: Automatic Mapping of OWL Ontologies into Java. *Proceedings of Sixteenth International Conference on Software Engineering and Knowledge Engineering (SEKE 2004)*, 20-24 June, Banff

⁸ Eberhart, A., 2002: Automatic Generation of Java/SQL Based Inference Engines from RDF Schema and RuleML. *International Semantic Web Conference (ISWC) 2002, Lecture Notes in Computer Science*, Springer-Verlag, Heidelberg, Volume 2342, pp. 102-116

⁹ Bernard, L., I. Kanellopoulos, A. Annoni, and P. Smits, 2005. The European Geoportal - one step towards the establishment of a European Spatial Data Infrastructure. *Computers, Environment and Urban Systems*, 29: 15-31.

¹⁰ Tait, M. G., 2005. Implementing Geoportals: applications of distributed GIS. *Computers, Environment and Urban Systems*, 29: 33-47.

¹¹ Bernard, L., U. Einspanier, S. Haubrock, S. Hubner, E. Klien, W. Kuhn, R. Lessing, M. Lutz, and U. Visser, 2004. *Ontology-Based Discovery and Retrieval of Geographic Information in Spatial Data Infrastructures*, Geotechnologien Science Report No. 4. http://www.delphi-imm.de/meanings/index_eng.html