Editor: James Hendler
University of Maryland
hendler@cs.umd.edu

Modeling with the Semantic Web in the Geosciences

Femke Reitsma, *University of Edinburgh*
Jochen Albrecht, *City University of New York*

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

This description makes evident the problems of both model interoperability and model simulator interoperability. Given the task's complexity and the number of research groups and individuals involved, there's a wide diversity of modeling approaches, such as models based on differential equations or stochastic methods. These approaches make difficult not only the interoperation of model specifications but also the intercomparison of models' structure and results, as is evident in the work of the Global Analysis, Integration, and Modeling Task Force (GAIM).[1] Similarly, in terms of simulator interoperability, models are developed in a broad range of programming languages and software, making it difficult to couple a Fortran model of thermohaline circulation with an ice sheet model in C++.

Compounding these concerns are spatial-data issues. Spatial data form a primary input for models, and, as with all other types of data, its volume continues to grow at an explosive rate.[2] Yet, worldwide the national clearinghouses for spatial data are experiencing a decline in use, management, and content owing to the community's dissatisfaction with the functional capability of the portals providing such data.[3]

Furthermore, from a computing perspective, much of the knowledge about modeled physical systems lies dormant in scientific papers, modeling code, and scientists' heads. Ontologies as knowledge repositories have been developed to support the primary goal of sharing knowledge in a manner that aids understanding.[4] However, the development of ontologies for geoscience disciplines has been limited to keyword lists for classification, such as the Global Change Master Directory's earth science keywords (http://gcmd.gsfc.nasa.gov/Resources/valids/keyword_list.html), or ontologies that are essentially class hierarchies with some limited expression of properties, such as NASA's SWEET (Semantic Web for Earth and Environmental Terminology) ontologies. Researchers have yet to tap the potential of ontologies and the Semantic Web for scientific modeling and simulation.

These problems largely derive from a common lack of explicit semantics in representing models,[5] spatial data, and scientific knowledge in general.

Process ontologies

To model earth system processes, we need ontologies in order to develop conceptually sound models, effectively communicate these models, enhance interoperability between models developed in different domains, and provide the opportunity for model components' reuse and sharing. To accomplish these goals, we must express these processes not only in terms of their types and properties but also 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 foundation of a process library that a simulation framework could use.

Modeling process behavior

We need an extended notion of ontology that can express rules defining the thresholds of process change and operations expressing the process's behavior. 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 initiate a set of typhoon processes, which in turn interact with other processes, such as coastal erosion. Thus, we need specifications that describe not only what the processes of the model are, but how those processes operate. Ultimately, this will support model components' interaction at the process-definitions level—that is, interoperation at the model components' level rather than interoperation as input and output to the model, which is the current dominant approach.

1541-1672/05/$20.00 © 2005 IEEE
Published by the IEEE Computer Society

DARPA is currently developing an extension to OWL called SWRL (Semantic Web Rule Language), which lets us express some aspects of rules and process behaviors. SWRL injects into OWL parts of RuleML (Rule Markup Language), thereby extending the set of OWL axioms to include Horn-like rules. So far, SWRL includes only 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 classes' properties or can call an external "oracle" with *BuiltIns*. SWRL BuiltIns have been developed for the language's future extensions, and are essentially calls to an external method or program that returns information required to evaluate the SWRL statement. We can use BuiltIns to incorporate programmed behavior by calling a program or implementation of some process behavior.

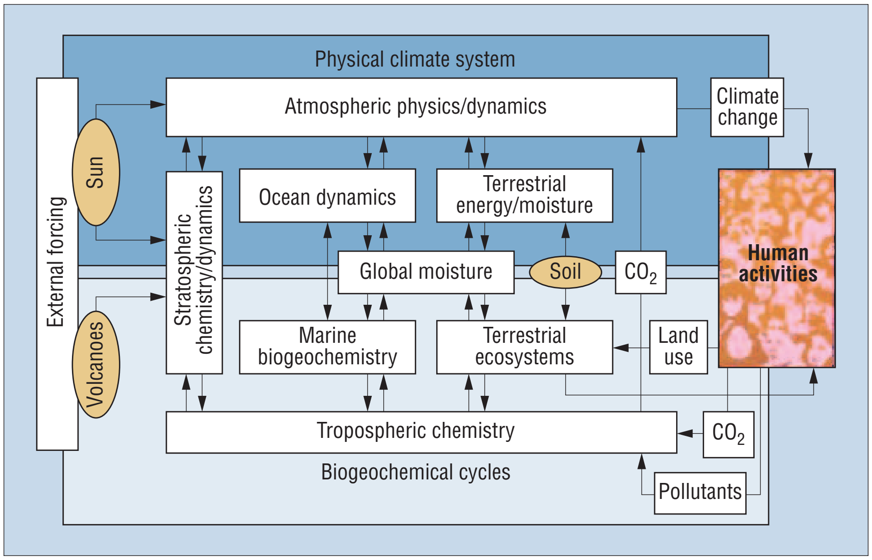

Figure 1. The Bretherton diagram of biospheric cycles. (figure courtesy of Earth Systems Science Overview, NASA, 1986)

Ontology-based simulation

Using an ontology that expresses all 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. Current research directed at converting Semantic Web languages to running code[6,7] makes such an ontology seem 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 in the future, this will lead to simulation platforms that assist in determining whether the process description's logic is correct according to scientific knowledge bases.

The Semantic Web will further facilitate researchers' collaboration and model components' automated discovery and use. Modeling the earth as a system requires an enormous breadth of knowledge of physical processes, a knowledge base that no individual scientist holds. Identifying what processes are important to model in certain systems requires the combined knowledge of an 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 might be peppered with process definitions that scientists have logged on their Web sites for automated modeling agents to discover and use.

Using spatial-data semantics

Spatial data plays a key role in modeling the earth system as the input to models and as a measure against which results are validated. However, spatial data's metadata is generally in a poor state due to disinterest in creating it when data is collected or modified. The lack of metadata might limit discovery of spatial data for modeling, but even with the full expression of metadata, the way this information is expressed limits the search for spatial data to predefined keywords—a search that is further hampered by the approach to presenting that data online.

Gazetteers, such as the World Wide Gazetteer or the Alexandria Digital Library Project's Gazetteer service, are the most common approach to geographic information retrieval. They let us search for geographic features such as cities, deserts, and jungles, or other information's location on the basis of the spatial location or attributes of those features. Implementations of gazetteers or gazetteer-centered search engines are based on 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 software's capabilities 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 agents) resulting in a significant amount of time spent learning the provider's structure.

Such spatial data providers typically adhere to syntactic metadata standards, such as the US federally mandated Federal Geographic Data Committee's Content Standard on Digital Geospatial Metadata and the ISO 19115 Geographic Information Metadata standard, and increasingly syntactic spatial-data standards, most notably GML (Geographic Markup Language) 3.0. More recently, these standards' semantics have been expressed in a set of ontologies produced at Drexel University. Yet, no accepted semantic standards exist for expressing the formal semantics of spatial data or the metadata describing that data.[8]

Rather than a geoportal being a Web site where geographic content can just be discovered,[9] the solution is ontology-based discovery and retrieval of geographic information.[10] However, beyond the expression of metadata's semantics for enhancing data discovery and spatial data use, 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, consider a user interested in finding spatial data that has a climate station within a specific watershed. The information to answer this query is expressed in the data's attribute tables and as a relationship between two different data sets, which isn't available in spatial data portals' metadata or functionality. Providing access to this information and its semantics will also enhance the use of model results for automated analysis and use.

Toward modeling the earth on the Semantic Grid

Beyond meeting the specific needs of modeling the earth and providing semantic solutions for interoperating among models and simulators, understanding of that system must be expressed and stored in a knowledge base. The lack of this knowledge base in a computable form will be the bottleneck for using Semantic Web technology for scientific research. Many recent, large-scale initiatives partly recognized this potential bottleneck**,** such as the SEEK (Science Environment for Ecological Knowledge) project and the GEOsciences Network project. These initiatives intend to create ontologies for their specific domains and use 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 using semantics in modeling the earth will enhance our ability to do science and could lead to new insights into the studied environmental systems and a greater understanding of methods used to represent those systems. The Semantic Web provides the platform for developing the solutions we've described and provides the opportunity to use the models and modeled results in new and interesting ways. For example, the challenges of natural disasters, such as the recent tsunami in Asia, highlight the potential utility of integrating models with spatial decision support systems for timely response to extreme events.

Related Links on Modeling Systems

Alexandria Digital Library Project: www.alexandria.ucsb.edu/gazetteer
Federal Geographic Data Committee standard:
http://fgdc.er.usgs.gov/metadata/contstan.html
Global Analysis, Integration, and Modelling:
http://gaim.unh.edu/about_gaim.html
Global Change Master Directory: http://gcmd.gsfc.nasa.gov
Geosciences Network project: www.geongrid.org
Drexel University's geospatial ontologies:
http://loki.cae.drexel.edu/~wbs/ontology/list.htm
OpenGIS Geography Markup Language: www.opengis.org/docs/02-023r4.pdf
ISO 19115 Geographic information, Metadata:
www.isotc211.org/scope.htm#19115
RuleML: www.ruleml.org
SEEK (Science Environment for Ecological Knowledge):
http://seek.ecoinformatics.org
SWEET (Semantic Web for Earth and Environmental Terminology) Ontologies:
http://sweet.jpl.nasa.gov/ontology
Semantic Web Rule Language: www.daml.org/rules/proposal
World Wide Gazetteer: www.gazeteer.com

References

1. J.C. Dutay et al., "Evaluation of Ocean Model Ventilation with CFC-11: Comparison of 13 Global Ocean Models," *Ocean Modelling*, vol. 4, no. 2, 2002, pp. 89–120.
2. C.T. Lu et al., "Two Web-based Spatial Data Visualization and Mining Systems: Mapcube & Mapview," Int'l Workshop Next Generation Geospatial Information, 2003, http://dipa.spatial.maine.edu/NG2I03/CD_Contents/EA/Lu_ChangTien_01.pdf.**//au:publisher? pages?//**
3. J. Crompvoets et al., "Assessing the Worldwide Developments of National Spatial Data Clearinghouses," *Int'l J. Geographical Information Science*, vol. 18, no. 7, 2004, pp. 665–689.
4. T.R. Gruber, "A Translation Approach to Portable Ontologies," *Knowledge Acquisition*, vol. 5, no. 2, 1993, pp. 199–220.
5. D.S. Mackay, "Semantic Integration of Environmental Models for Application to Global Information Systems and Decision-Making," *SIGMOD Record*, vol. 28, no. 1, 1999, pp. 13–19.
6. A.D. Kalyanpur et al., "Automatic Mapping of OWL Ontologies into Java," *Proc. 16th Int'l Conf. Software Eng. and Knowledge Eng.* (SEKE 2004), 2004, pp. 98–103. **//au: publisher? pages correct?//**
7. A. Eberhart, "Automatic Generation of Java/SQL Based Inference Engines from RDF Schema and RuleML," *Proc. 1st Int'l Semantic Web Conference* (ISWC 02), LNCS 2342, Springer-Verlag, 2002, pp. 102–116.
8. L. Bernard et al., "The European Geoportal: One Step Towards the Establishment of a European Spatial Data Infrastructure," *Computers, Environment and Urban Systems*, vol. 29, no. 1, 2005, pp. 15–31.
9. M.G. Tait, "Implementing Geoportals: Applications of Distributed GIS," *Computers, Environment and Urban Systems*, vol. 29, no. 1**,** 2005, pp. 33–47.
10. L. Bernard et al., "Ontology-Based Discovery and Retrieval of Geographic Information in Spatial Data Infrastructures," *Geotechnologien Science Report No. 4*, 2004, www.delphi-imm.de/meanings/index_eng.html.

Femke Reitsma is a lecturer at the Institute of Geography, School of Geosciences, at the University of Edinburgh. Contact her at femke.reitsma@ed.ac.uk.

Jochen Albrecht is an associate professor at the Department of Geography, Hunter College, City University of New York. Contact him at jochen@hunter.cuny.edu.