

The Open Geospatial Consortium and EarthCube

An EarthCube Technology Paper

Prepared by David Maidment¹, Ben Domenico², Alastair Gemmell³, Kerstin Lehnert⁴, David Tarboton⁵ and Ilya Zaslavsky⁶

October 2011

Introduction

EarthCube aims to create an integrated system to access, analyze and share information that is used by the entire geosciences community. The Open Geospatial Consortium (OGC) is an organization of which more than four hundred companies and agencies are members, whose purpose is to create open source standards for sharing geospatial and observational information. The authors of this paper are users of OGC standards in our work in hydrology, meteorology, climatology, oceanography and in the solid earth sciences, in other words, in key disciplinary fields that contribute to the Geosciences. Moreover, the OGC has an effective process for engaging individuals from many countries in standards development and interoperability testing that encourages them to set aside the roles they play in their home organizations and instead focus just on what is needed to share a particular class of information across the globe. This paper examines the hypothesis that an OGC-like process could be useful for broader sharing of information in the geosciences.

OGC Information Standards

The OGC has developed standards for sharing geospatial and observational information as web services. A web service is a function that enables one computer to make requests of another computer in a structured way and to receive appropriately coded responses in return. Sharing geoscience information requires the capacity to publish, catalog, discover, access and process information located on computers at any geographic location. OGC geospatial information service standards include:

- *Web Map Service* – for communicating georeferenced images;
- *Web Feature Service* – for communicating geospatial features as points, lines or areas with attached tabular attributes, that is, discrete spatial objects;
- *Web Coverage Service* – for communicating continuous spatial fields as multidimensional arrays;
- *Web Processing Service* – for executing functions or models on remote computers.

The OGC information framework for observational data is founded on an *observations and measurements model* that has four key components – a *feature of interest* has *properties* that are

¹ Center for Research in Water Resources, University of Texas at Austin

² Unidata Program Center, Universities Corporation for Atmospheric Research, Boulder, Colorado

³ Environmental Systems Science Centre, University of Reading, United Kingdom

⁴ Lamont-Doherty Earth Observatory, Columbia University, New York

⁵ Utah Water Research Laboratory, Utah State University, Logan, Utah

⁶ San Diego Supercomputer Center, University of California, San Diego

observed by a *process* that produces a *result*. This model originated in the solid earth sciences where it is used to describe properties of geological units as features of interest via sampling features such as transects, wells, or specimens. It has been generalized into the fluid domains where the feature of interest is understood to be a fluid domain in the vicinity of an observation instrument whose properties are being measured by that instrument. *Sensor Web Enablement* is a set of OGC specifications that define information encoding and service interfaces for interacting with a network of sensors through internet, and within this there is a *Sensor Observation Service* that describes how to query and obtain information from a particular instrument.

An organization may provide access to a large number of web services. To index them, OGC has developed a *Catalog Services for the Web*, which provides a single end-point web address through which the underlying services can be individually discovered and accessed, rather like going to the card catalog when entering a library in the traditional style.

Services in Solid Earth Science

In the solid earth sciences, a wide array of geospatially referenced data types is generated that include data from large sensor-based time series measurements to highly diverse and unstructured data generated ‘manually’ by investigators in the field or ex-situ in the lab. The integration of these diverse data types is critical for many of the large scientific questions and challenges in the Solid Earth Sciences. For example, major NSF-funded science initiatives such as Ridge2000, GeoPrisms, and the Critical Zone Observatories require analysis and integration of multi-disciplinary datasets. OGC-compliant web services can advertise and make accessible a wide array of useful geoscience data including station locations, physical specimens, event catalogs, and track lines. Several Web Map Service and Web Feature Service providers have already been successfully deployed in the solid earth science community, partly through collaborative work between UNAVCO, IRIS, the Marine Geoscience Data System, the National Geophysical Data Center, the Petrological Database of the Ocean Floor, and the LDEO Borehole Research Group, whose data are shown collectively in Figure 1.

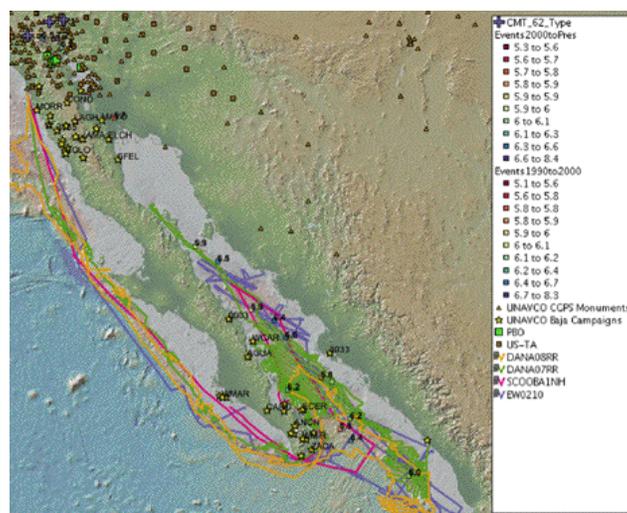


Figure 1. WMS and WFS features and layers include GPS campaign surveys and continuous stations, seismic stations, earthquake locations, seafloor bathymetry and expedition tracks.

Several OGC Observations and Measurements profiling efforts are emerging in the solid earth sciences, including EarthChemXML and SoilML. These will complement existing standards such as WaterML2 for hydrologic data, and the Climate Science Modeling Language (CSML) for atmospheric and oceanographic data.

Services in Atmospheric Science

Standards play a key role in allowing the Unidata Program Center to make data available to other communities without diluting the support provided to the core community of researchers and educators in the atmospheric sciences. While it is not practical for the Unidata Program Center to support the specialized analysis and display tools of other disciplines, it is important to make the datasets available via standard protocols so users from these other disciplines can continue to make use of their own tailored analysis and display tools with data available from Unidata. Standards are also enabling Unidata to work with others who provide complementary functions.

Web Processing Services are emerging as a key element in distributed systems that enable more analysis and display functions to be performed on the server side. This has several advantages, for example, in the era of "big data," the processing can be done where the data reside. Web Processing Services also makes it feasible to initiate and control sophisticated interactive data analysis and processing from mobile platforms. To gain a sense of power afforded by incorporating weather data into analyses done using tools from the GIS community, Figure 2 shows demographic information as an overlay on real-time, high resolution, local forecast models being run in regions of high precipitation probability. The weather data is conveyed in netCDF (network Common Data Form) via the OGC Web Coverage Service protocol. NetCDF is the first binary encoding format standard of the OGC.

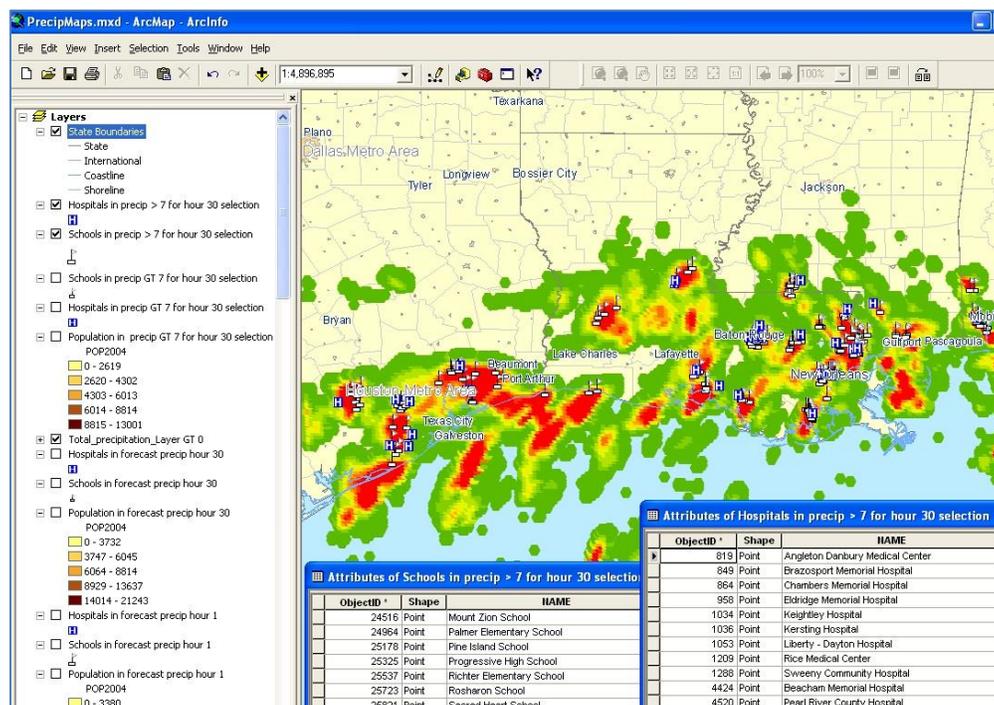


Figure 2. ArcMap Rendering of Schools and Hospitals in a Region of High Precipitation

Services in Ocean Science

Over the last five to ten years in the ocean sciences there has been an increasingly strong adoption of OGC standards, and related applications and projects. Many of these have employed the OGC Web Map Service standard and used it for serving gridded ocean model and satellite-derived data products. Examples of such projects and applications include the Integrated Ocean Observing System – OpenIOOS, MyOcean (which employs the University of Reading’s ncWMS software), IMOS – the Australian Integrated Marine Observing System and the SeaDataNet project. More recently there has been an upsurge in the use of other OGC web service standards, including extensive use of the the OGC Sensor Web suite of standards, especially the Sensor Observation service (SOS). This is becoming the preferred way of serving in situ data streams in the marine community, as opposed to the Web Feature Service which is not as well suited to handling time series data. Projects and initiatives which have incorporated the use of the SOS standard include OOSTethys, OpenIOOS and the OGC Oceans Science Interoperability Experiment.

Figure 3 shows a Web Processing Service by which the difference between two sea surface temperature grids is computed – the request for the computation is initiated at the University of Reading, the calculation is done with the grids in netCDF format at the British Atmospheric Data Center, and the result is delivered back to the University of Reading as a Web Map Service.

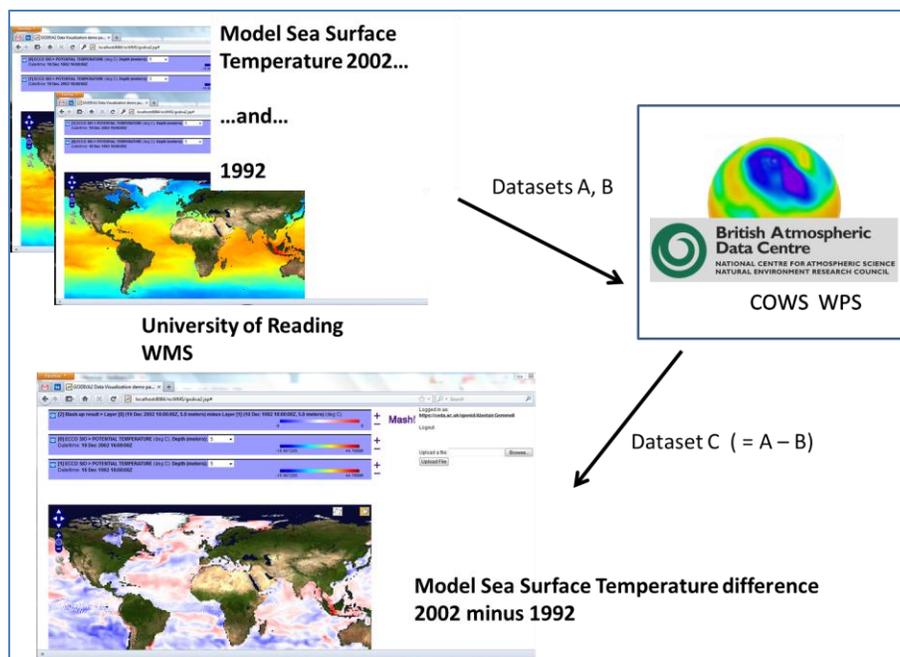


Figure 3. Sea Surface Temperatures – a Web Processing Service delivering a Web Map Service

Services in Hydrologic Science

Hydrologic science relies on combining information from multiple lines of inquiry. Figure 4 (a) shows a hydrologic scientist accessing information from GIS, modeling, observations, weather and climate modeling and remote sensing. Observations data measured at point locations are particularly important for hydrologic science, such as streamflow and rainfall gages, water

quality sampling points, and groundwater wells. Some of this information is obtained from intensely measured small watersheds, including the Critical Zone Observatories, but synoptic scale studies of the nation or large regions rely very largely on observations networks operated by water agencies within federal, state and local government.

The Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI) has invented a language, WaterML, for the communication of water observations data as a web service, and the US Geological Survey and some other water agencies have adopted WaterML and now publish their real-time and historical time series information in this format. CUAHSI has indexed water observations information at about 2.3 million locations in the US (Figure 4(b)), and provided access to this information in WaterML. Beginning in 2008, CUAHSI has been working with the OGC to modify WaterML, and a revised version, WaterML2, is being considered for adoption by the OGC as a profile or specialized version of the OGC Observations and Measurements specification, and thus as an international standard for encoding water observations data. In 2009, the Secretary General of the World Meteorological Organization and the President of the Open Geospatial Consortium signed an agreement to jointly develop internet information sharing standards for hydrology, meteorology, oceanography and climatology.

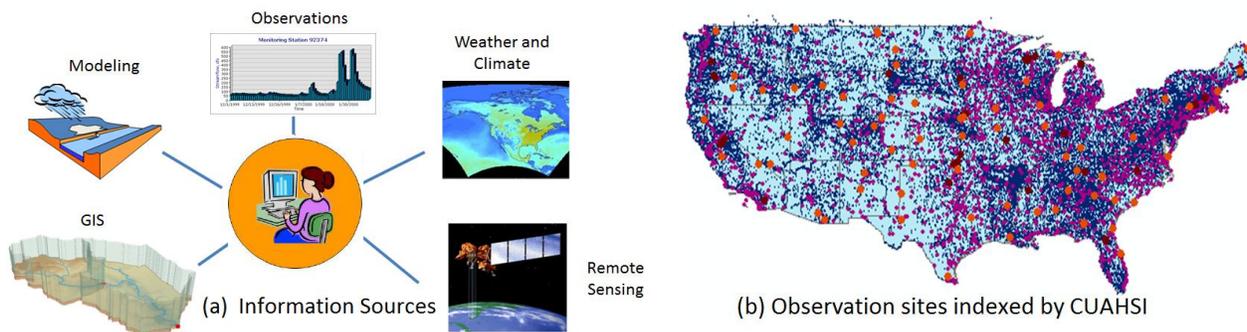


Figure 4. CUAHSI Hydrologic Information System

Geoscience Information Services

Any information system that emerges from EarthCube will build on the information infrastructure that NSF presently supports in the Geosciences. This includes several long-established data centers, such as the Unidata in Boulder and the IRIS Center in Seattle, and some newer initiatives including the CUAHSI Hydrologic Information System program and the Integrated Earth Data Applications at the Lamont-Doherty Earth Observatory. Each of these activities focuses on the needs of a particular domain geoscience community and because domain sciences tend to view the same physical circumstances in different ways, and even using different language, it is likely that there will always be the need for variations in terminology for different domain communities. This is similar to the task that the OGC deals with in developing international standards in the face of different languages of the countries its members represent. At its core, however, information about the earth rests upon a physical model of the shape of the earth and of the evolution through time of the phenomena that take place on and within the surface of the earth, its atmosphere and oceans. This suggests that it is possible to develop a common set of Geoscience Information Services that the domain science communities could share.

In geoscience cyberinfrastructure projects, reliance on service-oriented architectures is now a common approach to organizing distributed nodes of information. Standardizing on a limited set of information models and services, adopting them as community standards, and making the services available via representative hubs would help resolve a range of common challenges in the earth sciences. This vision requires that the certain nodes in the EarthCube system are developed to support easy translation between disciplines. They would include services for translating structure, syntax and semantics between information models, services and vocabularies adopted by different domains, and they would publish metadata catalogs that can be federated with metadata catalogs in other domains, for a one-stop data discovery and interpretation. In addition to “semantic web” elements, these translations may be helped by scientists themselves who will be able to connect and explain their data via social networks. EarthCube is first and foremost about connecting people.

OGC Process

The OGC has two component programs, a Standards program and an Interoperability program, which develop standards and test them in practical situations. Standards development occurs by taking existing standards and developing more extensive specifications, or profiles, that serve particular purposes. Interoperability testing is carried out by interoperability experiments, pilot studies and testbeds, that represent successively increasing degrees of commitment and involvement by the participant organizations in testing proposed standards. Domain working groups for particular disciplines focus activity in specialized areas of interest. The momentum of this activity is maintained by Technical Committee meetings that are held each three months at some location in the world, where standards are adopted and information exchanged in person among groups that otherwise communicate electronically.

Conclusions

The geosciences community embraces both the solid earth and fluid earth sciences and considers observed and modeled information distributed over the globe. Web services provide a mechanism for one computer to request information from another in a structured way for a variety of information types. The Open Geospatial Consortium is an international standards organization for developing web services for geospatial and observational information, including standards for cataloging and processing that information, and has recently joined with the World Meteorological Organization to jointly develop standards for information in hydrology, meteorology, climatology and oceanography. OGC standards are already fairly widely used in the NSF geosciences community. All these things suggest that EarthCube should seriously consider as one option for its development an OGC-based services-oriented architecture for the geosciences.

Should NSF decide to proceed in that direction, this could involve the formation of domain working groups within individual geoscience disciplines, the assembly of a Technical Committee for the Geoscience Information Services that considers common information standards across the geosciences, regular meetings each three months of the Technical Committee at NSF data centers around the United States, the advancement of web services standards in areas where current standards are deficient, and the testing of services provision through pilot studies where formal testing of the effectiveness of delivery of information to the geoscience community is undertaken.