Toward an Ocean Observing System of Systems

Luis Bermudez
MBARI 7700 Sandholdt Road, Moss Landing, CA 95039

Philip Bogden
Southeastern Universities Research Association, 1201 New York Ave, NW, Washington DC, 2001 and GoMOOS, 350 Commercial St., Portland, ME 04101

Eric Bridger
GoMOOS 350 Commercial Street, Portland ME 04101

Gerry Creager
Texas A&M University 1700 Research Parkway Ste 160, College Station, TX, 77843

David Forrest
Virginia Institute of Marine Science, Route 1208, Greate Road, Gloucester Point, VA, 23062

John Graybeal
MBARI 7700 Sandholdt Road, Moss Landing, CA 95039

Abstract **The earth- and ocean-science communities are developing the concept of a "system of systems" (www.epa.gov/geoss) for observing the earth and oceans. Related initiatives in the ocean sciences range from the application-oriented Integrated Ocean Observing System (IOOS) to the research-oriented Ocean Observatories Initiative (OOI). In an ideal world, all ocean observations would support the broad range of activities because all the systems would be interoperable. Such a "system of systems" will surely result from standardization of some kind. One challenge is that we already have many standards that address data-encoding formats, content metadata, protocols for communicating between computers, and ontology languages for knowledge representation. Two grass-roots community initiatives have aligned to make some concrete choices that will advance the "system of systems" concept: the Marine Metadata Interoperability (MMI) interoperability demo (www.marinemetadata.org) and the OpenIOOS interoperability test bed (www.openioos.org). Although they have substantial overlap, each initiative brings a complimentary set of experiences to the table. With funding from NSF and NOAA, the MMI project has been enabling the exchange, integration and use of marine data by emphasizing ontologies that employ the Web Ontology Language (OWL), in anticipation of the Semantic Web. With funding from the Office of Naval Research and NOAA, OpenIOOS participants have been demonstrating that "standards enable innovation" by leveraging web-service and data-model specifications developed by the Open Geospatial Consortium (www.opengeospatial.org). We present the lessons learned from this coordinated and combined effort.**

I. INTRODUCTION

Two grass-roots community initiatives have aligned to make some concrete choices that will advance the "system of systems" concept: the Marine Metadata Interoperability (MMI) demonstration project (www.marinemetadata.org) and the OpenIOOS interoperability test bed (http://www.openioos.org). The alignment is based on agreements about how to share marine observed data. The agreements are based on concrete choices about content standard, protocols, interfaces and controlled vocabularies; and sharing a common Ocean Observing System (OOS) architecture, for the purpose of clarifying the functionality that must be developed.

II. MMI, SURA AND TETHYS

OpenIOOS.org has been demonstrating that "standards enable innovation" for several years, thanks to contributions from Ocean Observing Systems (OOSs) around the country. With support from the NOAA Coastal Services Center and the Office of Naval Research, the Southeastern Universities Research Association (SURA) has helped organize OpenIOOS.org as part of the SURA Coastal Ocean Observing and Prediction (SCOOP) program (http://scoop.sura.org).

SURA, through the SCOOP program, supports the vision of enabling a distributed laboratory for research and applications with broad participation from research institutions working in partnership with federal agencies and the private sector. Activities like OpenIOOS.org demonstrate important steps toward achieving communities of distributed scientists working together. So SURA has hosted an ongoing series of community-oriented "OOS Tech" workshops that have advanced the use of web services for interoperable ocean science. And OpenIOOS.org, which is a proof of concept portal interface to the distributed participants, demonstrates a Service Oriented Architecture (SOA) with heavy emphasis on Open Geospatial Consortium (OGC) web-service specifications.

The first step to creating OpenIOOS.org occurred in 2004, when GoMOOS (www.gomoos.org) and SEACOOS (www.seacoos.org) deployed the first two OGC-compliant real-time data services that were aggregated and visualized at OpenIOOS.org. Within only a few months, the SOA demo grew to include several more OOSs and over a dozen university, non-profit and federal agency partners, including NOAA, NASA and Navy. Remarkable aspects of the demo include the heterogeneity of the underlying data systems, the highly distributed nature of these systems, and the variety of technologies used to implement the services, which include a combination of Open Source (e.g., University of Minnesota's MapServer) and commercial (e.g., ESRI) software. Integration and interoperability were achieved quickly and economically thanks to broad adoption of the OGC interface specifications.

The Marine Metadata Interoperability project started in September 2004. After almost two years it comprises more than 300 members, with a web site consisting of thousands of pages with guides and references about marine metadata issues. One of the activities of MMI is an interoperability demonstration using a service-oriented architecture, common use content standards, and semantic mediation via ontologies. MMI produced an initial result integrating heterogeneous systems at MBARI, via a metadata interface using Dublin Core Metadata Initiative and SOAP web services. MMI has also created tools to facilitate the creation, mapping and publishing of ontologies. The semantic mediator is a web service that fits in a service-oriented architecture solving semantic issues among heterogeneous systems.

In 2005 MMI, OpenIOOS and SCOOP started working together on a single, combined interoperability test bed activity that combines their programmatic interests of having distributed services and adopting a service-oriented architecture to share real time data. The two interoperability demonstrations were merged in June 2006 under the name of Tethys. The combined team set up an initial set of metadata requirements such as geo-spatial location and platform type, and agreed on an interface to serve as a wrapper for each data system. Fig. 1 depicts some initial results. The interface is being transitioned to Sensor Web Enablement (SWE) technologies, leveraging the work of OGC.

Figure. 1 Latest Salinity Data from Tethys Data Providers. http://www.openioos.org/testbed/salinity/index.html

III. COMMUNITY AGREEMENTS

When two or more communities want to interoperate, they make decisions about the common protocols, interfaces, controlled vocabularies, and metadata standards. A brief description for each of these four items of agreement is given in this section, as well as the rationale behind the choice that MMI/SURA have undertaken.

A. Protocols

A protocol is a strategy for transmitting data between systems. Several protocols exist. Two widely used XML-based web-service implementations have special relevance for Tethys: SOAP/WSDL and REST. Implementations based on SOAP/WSDL tend to be easier to put together in a chain of processes or workflows. For example, Standards Organizations like OASIS (Organization for the Advancement of Structured Information Standards) are currently working on Web Services Business Process Execution Language (BPEL) that is based on SOAP/WSDL.

The types of OGC services used to implement the early versions of www.openioos.org are not SOAP/WSDL based; however there has been progress related at OGC relevant to the issue [1]. Our impression is that OGC specifications, standards and tools will increasingly support WSDL/SOAP protocols.

B. Standard Interfaces

When two information system want to exchange data in a client/server arrangement, they can exchange one or more XML messages, or they can simply access data through a URL, as with an FTP server or an OPeNDAP URL. OGC has defined XML-based interfaces that have been widely accepted by the community. The most widely used OGC interface is WMS (Web Map Service), which allows one to share maps on the web. WFS (Web Feature Service) is a data-exchange for geospatial "feature" data (e.g., point, lines etc.). The WCS (Web Coverage Service) interface facilitates exchange of geospatial "coverages" that included spatial and temporal grids data.. The GALEON [2] initiative with the OGC is investigating applications of WCS that leverage the OPeNDAP and THREDDS technologies that are widely used in the oceanographic and meteorology communities.

Some of the more recent OGC specifications are evolving from the Sensor Web Enablement (SWE) initiative, including SensorML and SOS (Sensor Observation Service). These standards are still evolving, and they specialized in dealing with sensor descriptions and sensor data. MMI and SURA are currently implementing SWE interfaces as part of Tethys. in an attempt to investigate the complementary relationships to the WFS and WMS Implementations that underlie OpenIOOS.org.

C. Controlled Vocabularies

A controlled vocabulary can be defined as a set of restricted words, used by an information community when describing resources or discovering data. Having controlled vocabularies harmonized, align and map is essential to achieve semantic interoperability among information communities [3]. Controlled vocabularies represented in ontology languages, like OWL [4] can leverage of all the tools for inferencing and discovery being developed for the Semantic Web [5].

In Tethys, each term like, platform type, measurement type, unit, etc. is given as a Unique Resource Identifier (URI) [6]. These terms are well defined in an Ontology hosted at MMI. The ontology is based on standardized controlled vocabularies like CF [7], GCMD [8] ensuring the maximum level of semantic interoperability, among other observing systems.

D. Metadata Specifications

A metadata specification is a set of statements that helps domain experts to formally express the rules of usage for metadata elements. A metadata element is for example, title, creator, latitude, longitude and sensor type. Rules are necessary to declare that an element is mandatory or that it can be repeated.

Various metadata specifications exist (see comprehensive list [9] [10]); however, in most cases but they are too general and so lack of domain-specific elements [10]. SensorML [11], an OGC specification, provides the capabilities to describe the inputs and output of a sensor. SensorML also has the flexibility for describing domain-specific properties and types via URIs for observed phenomena, platforms and sensors. Therefore, SensorML could serve as the basis for interchanging interoperable sensor data in the marine community.

IV. TETHYS FRAMEWORK

One definition of a system of systems considers a system that exists when subsystem elements are sufficiently interoperable with each other that they comprise the proverbial whole that is greater than the sum of its parts. We consider the necessary ingredients of an Ocean Observing System of Systems (OOSS) to be comprised of a Data Provider, a Registry, a Mediator, a Data Aggregator, a Data Analyzer and a Viewer. Each is part of an individual OOS, but the System of Systems itself is part of a larger collection of OOSs, each acting as a Data Provider for the parent OOSS. In the Russian culture, a painted, wooden, nested doll, the Matryoshka, provides a particularly colorful metaphor: A system, within a system, within a system (Figure. 2).

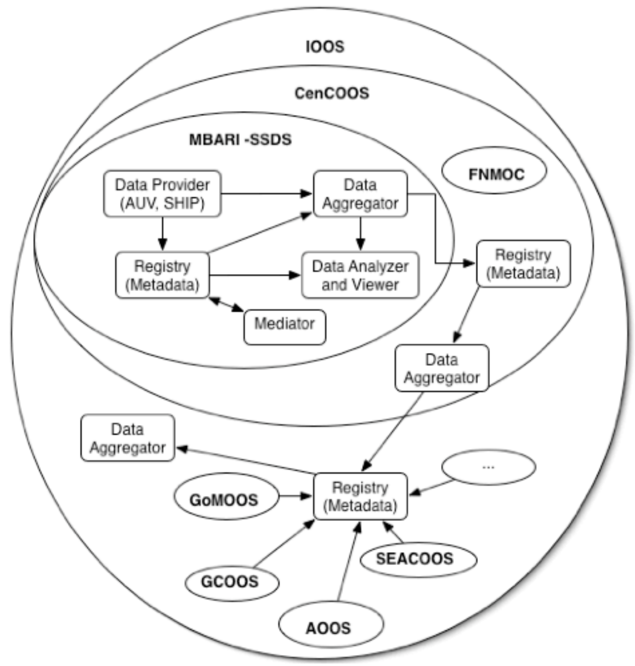

Figure. 2 Matryoshka and Ocean Observing Systems

In software systems development, breaking the pieces of a system into modular components helps to implement each component, specialized it and reuse it when necessary. MMI and SURA have agreed to start with the following working definitions of the components of a system as follows:

A. Data Provider

The Data Provider is responsible for collecting data from instruments (*in situ* sensors or samplers), manipulating the data (e.g., time series averaging, quality control processing), or running a numerical model (which itself produces “data”). The Data Provider is also responsible for storing his own data and registering the details of the services that it has made available. Services are defined in a service-contract, negotiated between cooperative entities with similar data and data sharing requirements. For example, the Data Provider may register a Web Service in order to get metadata in XML conforming to ISO 19139 Schemas (the implementation model of ISO - 19115).

B. Registry

Registries manage information about Data Providers. A

Registry allows a Data Provider to register, and facilitates other components of the architecture in performing queries of its contents. A registry might be a metadata catalogue service, a clearinghouse, a Universal Description, Discovery and Integration (UDDI) registry, a list of OPeNDAP Uniform Resource Locators (URLs) or a TREDDS catalogue. Ideally, a registry could engage in permanently discovering Data Providers, the same way as search engines crawl the World Wide Web (Web) registering the URLs and their content in their registry. A registry uses a mediator to help discover data.

C. *Mediator*

Mediators allow an OOS to discover data and to get the data in a useful format to a user. A semantic mediator would find a relation between the term "upward_sea_water_velocity" used in the CF [7] convention and "Ocean Currents", the term used by GCMD [8]. Therefore, when a user searches for "Ocean Currents", data sets containing "upward_sea_water_velocity" will be retrieved.

D. *Data Aggregator*

Data Aggregators acquire data from Data Providers. Aggregators derive necessary information about the Data Providers and their data via information maintained in the appropriate Registry. A Data Aggregator would know what language is used by each Data Provider, how to interpret the Data Provider's metadata, and how to access the data, and how to process it for special uses, including computational analysis and archival storage. Data Aggregators are utilized by other Data Aggregators to provide a data conduit to a bigger region, as well as by Data Analyzers and Viewers.

E. *Data Analyzer and Viewer*

Data Analyzers and Viewers may be Web sites such as openioos.org or stand-alone applications that provide the Graphical user Interface (GUI) to allow the user to view and analyze the data, such as the OGC-compliant WorldWind application from NASA. Others examples include the Live Access Server (LAS), Unidata's Integrated Data Viewer (IDV), Google Earth, etc. Analysis may involve user-selected integration of various datasets for simultaneous (layered or overlaid) display, or numerical or computational exercises which will yield a new resulting product for examination; many GIS analysis tools already behave as OGC compliant clients and servers (e.g., Minnesota MapServer).

V. CONCLUSIONS

The teams of scientists and software engineers from SURA and the MMI project that drive Tethys have every intention to freely and openly share all experiences that come out of this combined initiative. In this sense, they share a common "Open Source" attitude about engaging a broad array of participants in their vision for a system of systems. The bigger system should engage communities of participants from government, private sector and research communities. These three communities may be thought of as the three legs of a stool that supports a broad array of stakeholders. These stakeholders stand to benefit greatly from the concept of a system of interoperable observations. Many of the stakeholders who benefit from the system also find themselves in one of the three legs of the stool. The technologies described here enable the interactions that will make the system of systems possible. If you would like to participate, please contact one of the authors.

ACKNOWLEDGMENTS

This activity is supported by the National Science Foundation under award number ATM-0447031; Monterey Bay Aquarium Research Institute and the David and Lucile Packard Foundation; Office of Naval Research, Award N00014-04-1-0721; and, by NOAA Ocean Service Award NA04NOS4730254. Any Opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

REFERENCES

[1] J. Sonnet, C. Savage, OWS 1.2 SOAP Experiment Report, OGC, 2003.
[2] Galeon IE, http://www.unidata.ucar.edu/projects/THREDDS/GALEON/GALEON-Activity-Plan.htm
[3] L.E. Bermudez, J. Graybeal, A. Isenor, R. Lowry, D. Wright, Construction of Marine Vocabularies in the Marine Metadata Interoperability Project, Ocean's 2005, Washington, 2005, pp. 14-18.
[4] S. Bechhofer, F.v. Harmelen, J. Hendler, I. Horrocks, D.L. McGuinness, P. Patel-Schneider, L.A. Stein, OWL Web Ontology Language Reference, W3C Recommendation, 2004.
[5] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, Scientific American 184(2001) 34-43.
[6] T. Berners-Lee, F. R., L. Masinter, Uniform Resource Identifiers (URI): Generic Syntax, W3C/IETF Network Working Group, 1998.
[7] CF Standard Names, 2006, http://www.cgd.ucar.edu/cms/eaton/cf-metadata/index.html
[8] L.M. Olsen, G. Major, S. Leicester, K. Shein, J. Scialdone, H. Weir, S. Ritz, C. Solomon, M. Holland, R. Bilodeau, T. Northcutt, T. Vogel, Global Change Master Directory (GCMD) Earth Science Keywords, http://gcmd.gsfc.nasa.gov/Resources/valids/keyword_list.html
[9] MMI References for Content Standard and Profiles, http://marinemetadata.org/content/refs
[10] L.E. Bermudez, Ontomet: Ontology Metadata Framework, Ph.D. Thesis, Drexel University, 2004.
[11] SensorML, http://vast.nsstc.uah.edu/SensorML/