

MAKING THE MOST OF INFORMATION FROM ENVIRONMENTAL MONITORING SYSTEMS

Madilyn Fletcher, Dwayne Porter, Jeremy Cothran, Dan Ramage

AUTHORS: Madilyn Fletcher, School of the Environment, University of South Carolina, Columbia, SC; Dwayne Porter, Baruch Institute for Marine and Coastal Sciences & Dept. of Environmental Health Sciences, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208; Jeremy Cothran and Dan Ramage, School of the Environment, & Baruch Institute, University of South Carolina, Columbia, SC 29208;

REFERENCE: *Proceedings of the 2010 South Carolina Water Resources Conference*, held October 13-14 at the Columbia Metropolitan Convention Center

Abstract. The University of South Carolina has a long-standing and extensive effort in the development and implementation of a range of platforms and tools enabling interoperability among distributed data monitoring activities in southeastern U.S. coastal regions. These activities are central to a number of ocean observation programs, including the Carolinas Regional Coastal Ocean Observing System (Carolinas RCOOS), the Southeast Coastal Ocean Observing Regional Association (SECOORA), and the Center for Integrated Information System Testbed Project. These programs were preceded and made possible by the earlier Carolinas Coastal Ocean Observing and Prediction System (Caro-COOPS), Coastal Ocean Research and Monitoring Program (CORMP), and Southeast Atlantic Coastal Ocean Observing System (SEACOOS) collaborative programs. All of these initiatives involve a range of real time and delayed mode data on environmental conditions, which are received and managed to support a large variety of data uses and products. The capacities that have been established during the development of these programs involve the following: (1) Ability to access a variety of data types and import them into sophisticated centralized data hubs; (2) A relational data base structure for rapid access and selection of the data; (3) A system of quality assurance and quality control that helps to identify spurious data and/or interesting data anomalies; (4) A variety of platforms and tools, including map-based products that enable visualization presentations of data/information; (5) Tools that allow users to pull data of interest and that push data to users, according to user-specified needs; (6) An understanding, and initial implementation, of the infrastructure required for reliable data access, maintenance, and supply as needed. These processes and protocols are largely based on open source software and involve a variety of partners that add diversity, resources, and skills to the collective initiative. The potential to expand this to a variety of environmental monitoring systems is enormous. A wide

range of federal and state government, university, and private sector programs collect enormous amounts of environmental data for a range of purposes, such as weather forecasts, beach contamination monitoring, and fisheries resource management. While each of these observing efforts serves a specific purpose, a capacity to integrate and access such disparate data streams would increase their value and application potential enormously.

INTRODUCTION

To understand the dynamics of our environmental systems, such as coastal oceans, rivers, and groundwater aquifers, it is necessary to monitor important measurable conditions over time, develop an understanding of natural variability, and establish a baseline against which change can be detected and measured. Large and small monitoring programs have been established for various specific purposes, e.g. weather forecasting, coastal ocean water and atmospheric conditions, water quality assessment. In each case, data management systems have been developed to ensure that monitored data were organized in a manner that optimized their use and accessibility. However, in many, if not most, cases where large databases must be continually developed and managed, the data management infrastructure has been designed specifically for the purpose at hand, which often limits ready access to those data for alternative uses. Increasingly, environmental monitoring programs have recognized the importance of establishing “interoperability” among their data management systems by adoption of a limited set of standards, processes, and protocols and/or by developing middleware that facilitates transfer of information between various users. An excellent example of this move towards interoperability is represented by a range of coastal ocean observation and monitoring systems and the NOAA

Integrated Ocean Observing System, IOOS[®] (www.ioos.gov; NOAA, 2009).

BACKGROUND AND RELATED WORK

The University of South Carolina (USC) has over the past decade established a range of platforms and tools that foster interoperability among a variety of distributed data monitoring activities in southeastern U.S. coastal regions. These activities are central to a number of ocean observation programs, including the Carolinas Regional Coastal Ocean Observing System (Carolinas RCOOS; www.carolinasrcoos.org), the Southeast Coastal Ocean Observing Regional Association (SECOORA; www.secoora.org), and the Center for Integrated Information System Testbed Project (Fletcher et al., 2009). These programs were preceded and made possible by the earlier Carolinas Coastal Ocean Observing and Prediction System (Caro-COOPS; www.carocoops.org), Coastal Ocean Research and Monitoring Program (CORMP; www.corp.org), and Southeast Atlantic Coastal Ocean Observing System (SEACOOS; www.seacoos.org; Fletcher et al., 2008) collaborative programs. These programs also provide data for incorporation into national data systems, such as the NOAA National Data Buoy Center (NDBC), and in more recent years, they have established linkages and interoperability with the NOAA National Estuarine Research Reserve System (NERRS). All these programs involve a range of real time and delayed mode data on environmental conditions, which are accessed and managed to support a large variety of data uses and products.

PROGRAM DESIGN

Critical functions of these coastal ocean data management systems include the ability to aggregate large volumes of distributed data, the ability to search for and access subsets of data, the provision of data through a variety of formats or “services,” and a mechanism for assessing metadata or quality of subsets of data.

Aggregation of Data

An initial need was to be able to access data from multiple sources and aggregate them in a way that enabled various combinations of data to be collectively used for additional purposes. The various data sources tended to be geographically distant and based on processes and protocols that had been independently adopted, so that a variety of database formats and protocols were encountered. We determined that a relational database format that was oriented towards the

type of observation, as opposed to geographic location for example, was most flexible and efficient. Next steps were to design the database schema so that it enabled data access from multiple sources, supported data querying by users, and supported continued growth of the database as monitored data accreted and new data providers or observation platforms were added. The resultant simple database schema, named Xenia (<http://code.google.com/p/xenia/wiki/XeniaHome>), is well suited for rapid processing of real time data, and is currently serving approximately 5,000 hourly observations from 1,000 separate platforms. The data are indexed using the observation time and type, and the geographic location of the observations and platforms are stored in the database, allowing the use of OpenGIS (<http://www.opengeospatial.org/>) geospatial queries on the data as well. Although well suited for speed, one of Xenia’s limitations is that it is not suitable for large volume images or raster data sets. However, metadata on such large volume databases can be included in the Xenia database, facilitating their location by potential users.

The broad utility of Xenia has been demonstrated by its adoption, or leverage, by a number of other observing programs. These include:

- The National Weather Service Marine Weather Portal(MWP) which merges the latest regional observations with other NWS-specific products such as storm advisories or hazards (<http://forecast.weather.gov/mwp>)
- The Carolinas Regional Coastal Ocean Observing System (<http://carolinasrcoos.org>)
- The Great Lakes Information Network (<http://gis.glin.net>)
- Northwest Association of Networked Ocean Observing Systems (NANOOS) (<http://code.google.com/p/xenia/wiki/NANOOSXenia>)
- The Intelligent River hydrology application (<http://www.intelligentriver.org/data-products>)

Data Discovery

To optimize the utility of an extensive database, it is important that potential users can discover desired information through readily accessible, simple search mechanisms. Thus, a number of registries or catalogues are being developed to document available databases and compensate for differences in their original properties. For example, a common difficulty is the use of different “vocabularies” by data collectors as they document their data. A significant effort to deal with this problem is the Marine Metadata Initiative (MMI; <http://code.google.com/p/mmisw/> <http://marinemetadata.org>), which has been addressing the problem of “semantic interoperability” to facilitate data exchange. The MMI approach has been to design a framework that includes

operationally tested tools and community-based procedures that are suitable for current needs and can progress towards an inclusive, interoperable semantic infrastructure. The framework includes a vocabulary builder, ontology registry, and semantic engine, which link the data providers and a variety of users.

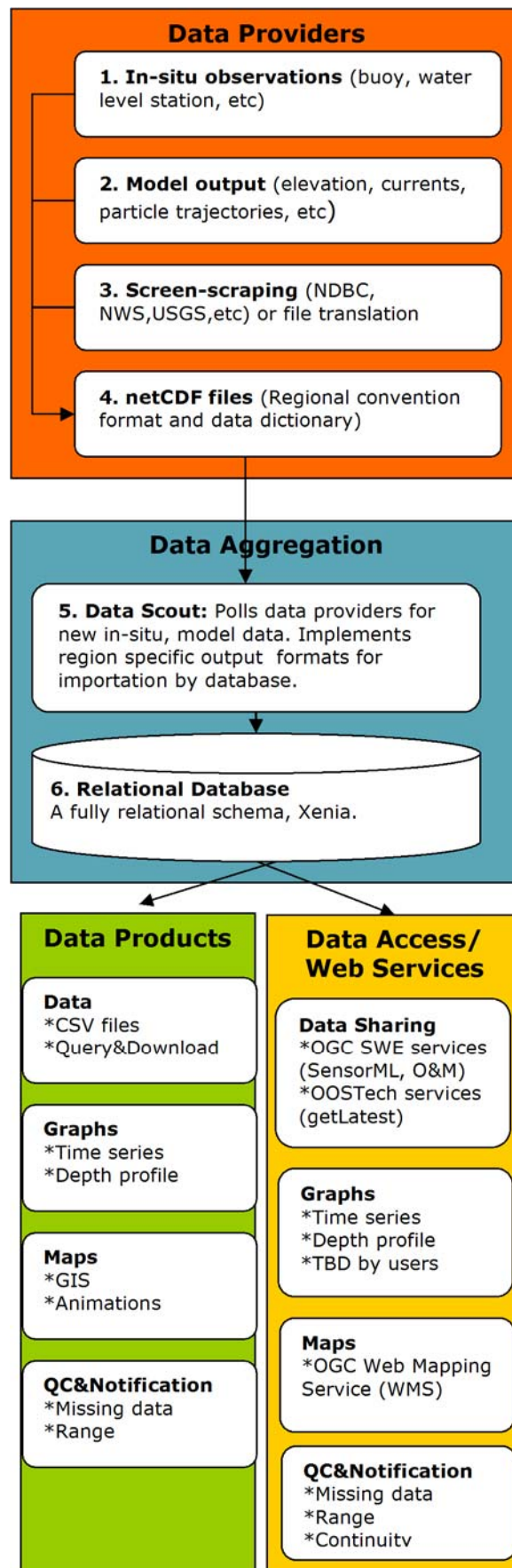
Other registries that are working to facilitate interoperability and data exchange include the National Observation Registry (<http://obsregistry.org>); SECOORA sensor inventory and user registry; and EcoWatch, developed by the NOAA National Coastal Data Development Center (NCDDC; <http://ecowatch.ncddc.noaa.gov/>) . For CarolinasRCOOS, we have developed an online spreadsheet (<https://spreadsheets.google.com/ccc?key=0AkGdvmaCDgbOcjZpSTU4XzBOem0wQ3RKZi0xZUF0T1E&hl=en>) and browser catalog (<http://rcoos.org/carolinasrcoosrev2/dataLinks.html>), and a listing of the existing data files, maps, and services are available. Our objective is to use the data from this spreadsheet to help cross-populate other applications or inform users of data links.

Enhancing Data Availability

The ultimate purpose of amassing large volumes of data is to be able to optimize their use for various assessment or analytical purposes. Thus, the database serves as a central organized platform from which data can be exported via a variety of formats and services. Much of our recent focus has been on identification of various data applications and designing the services that provide new applications and data utility. Accordingly, we have developed code for a variety of formats that are periodically delivered to users and for data-based “services,” which are custom-designed for particular types of users. Further information on these formats and services are at <http://code.google.com/p/xenia/wiki/XeniaProducts>. Examples include the following:

- Google Maps/Earth, Latest ObsKML, Styled KML – allows mapping and sharing of KML/KMZ data via Google Maps/Earth
- Time-series graphs – individual sensor/ observation graphs
- GeoRSS – platform-specific GeoRSS feeds which provide hourly updates for RSS type readers
- GeoJSON – platform-specific GeoJSON feeds which provide hourly updates for Javascript oriented browser functionality

Figure 1. Schematic diagram illustrating data flow to various products and services. Further information can be accessed at http://secoora.org/data/data_feeds .



- HTML tables – platform-specific HTML tables which can be dynamically pulled for AJAX style applications doing webpage HTML content swapping
- SQL,CSV,Shapefile – SQL for same schema Xenia databases, CSV for Excel or general import, Shapefile for GIS style applications
- Open Geospatial Consortium (OGC) web services such as WMS(Web Mapping Service), WFS(Web Feature Service) and SOS(Sensor Observing Service)
- Quality-control and event notification – the archival collection of data allows for self-referential quality-control and trending of data and also event notification when observations exceed known historical parameters.

METHODS

The fundamental philosophy underlying the development of this interoperable data management infrastructure has been incorporation of shared problem-solving and collaborative development. The University of South Carolina, the University of North Carolina at Wilmington, University of North Carolina at Chapel Hill, and the University of South Florida have been key participants in the development of ocean and coastal observing programs in the Southeast, and have played essential roles in the development of the data management infrastructure. Communication and information exchange is a continual practice, and there is ready sharing of technical tools or processes used by the programs. Progress is discussed and documented by online wikis, white papers, and other communication platforms (see www.secoora.org). To the extent possible, open-source tools are used for development of new capacities, which enables others to more easily share, combine, and reuse their experience and development work with others.

CONCLUSIONS

The establishment and evolution of coastal ocean observing programs and IOOS® initiatives has provided a vehicle for developing an understanding of the issues, needs, and mechanisms to achieve data interoperability. A variety of approaches and tools have been developed and tested to verify their ability to promote interoperability of diverse types of real time and delayed mode databases. Yet there are large numbers of environmental databases that are largely confined to their original purposes because of original design and

practices. The value of these data and their capacity for problem-solving can be enhanced enormously when steps are taken to overcome semantic and/or platform interoperability obstacles. We need to address the need for interoperability by (1) identifying those databases sources that can be made interoperable through techniques already in hand, thus enhancing their utility within a broader environmental framework, and (2) before embarking on new data monitoring programs, identify the database infrastructure options available so that decisions are made that promote the interoperability of the newly acquired data, rather than limiting their access to the relatively limited users associated with the program.

ACKNOWLEDGEMENTS

This work has evolved over a number of years with support from NOAA grants numbers NA16RP2543, NA07NOS4730222, NA07NOS4730222, NA08NOS4000476, NA08NOS4730409 / S453, and US ONR grant number N00014-02-0972. The contributions of a large variety of participants in SECOORA, Carolinas RCOOS, Caro-COOPS, CORMP, and SEACOOS have been essential foundations of this work.

LITERATURE CITED

- Fletcher, M., J. Cleary, J. Cothran, and D. Porter. 2008. Southeast Atlantic Coastal Ocean Observation System (SEACOOS) Information Management: Evolution of a Distributed Community System. *Marine Technology Society Journal*. 42:28-34.
- M. Fletcher, J.R. Pournelle, D. Ramage, D.E. Porter, V. Shervette, R.H. Kelsey. 2009. A Southeast Regional Testbed for Integrating Complex Coastal and Ocean Information Systems. Proc. of the MTS/IEEE Oceans 09 Conference. Biloxi, MS. Oct.26-29, 2009.
- NOAA, 2009. Integrated Ocean Observing System (IOOS®) Data Management and Communications Concept of Operations, Version 1.5, http://ioos.gov/library/dmac_cops_v1_5_01_09_09.pdf.