INTEGRATION OF ENVIRONMENTAL INFORMATION SYSTEMS

Madilyn Fletcher, Dwayne Porter, Jeremy Cothran, Jesse Cleary

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, School of the Environment, Baruch Institute, & Advanced Solutions Group, University of South Carolina, Columbia, SC 29208; Jesse Cleary, Department of Marine Sciences, University of North Carolina, Chapel Hill, NC, 27599
REFERENCE: *Proceedings of the 2008 South Carolina Water Resources Conference*, held October 14-15, 2008, at the Charleston Area Event Center.

Abstract. The University of South Carolina (USC) has been involved in developing and implementing a range of information management efforts aimed to create interoperability among distributed data monitoring activities in southeastern U.S. coastal regions. A major example of such programs is the Southeast Atlantic Coastal Ocean Observing System (SEACOOS), led by the University of North Carolina, with primary partners at USC, University of South Florida, Skidaway Oceanographic Institution, and the University of Miami. SEACOOS has been incorporating real-time monitoring observations from a region spanning the west Florida shelf to the NC-VA border, encompassing a variety of environmental variables and monitoring technologies. SEACOOS established a comprehensive information management (IM) system, incorporating distributed hardware and software applications at the partner institutions and establishing IM hubs for data aggregation and data product dissemination. This extensive system can serve as a model for distributed observations of other water resources. The data remained largely distributed at the participating institutions, and, to the extent possible, we used existing resources and practices, so that the participating institutions would maintain their flexibility and autonomy, resources would be optimized, and implementation of the IM system could proceed as rapidly as possible. All data and information applications are freely available, usually via a web browser. An important initial step was development of a set of standards for data transport mechanisms, vocabulary, and metadata. A relational database structure was adopted and adapted as we identified needs, e.g. improved speed in aggregating distributed data and posting it to the web site. Considerable focus has been placed on the development of map-based products by aggregating the distributed data and developing a range of potential maps with selectable variables (information layers). To facilitate access to map-based products and optimize server use, a range of pre-generated maps, which refresh every hour, have also been generated. Map development was based on parallel developments in database structure, data standards, and data transport mechanisms. One of our most important considerations has been to develop an IM infrastructure that could be expanded beyond its initial application in the southeast coastal region to other U.S. coastal regions and could be interfaced with the broad range of environmental data generated in the watersheds. The "lessons learned" within SEACOOS have instructed us on key processes required to achieve interoperability and infrastructure requirements for stability and appropriate redundancy.

INTRODUCTION

One of our greatest challenges in the "information age" is to be able to share and integrate information that resides in different electronic infrastructures. Such "interoperability" among disparate data and information systems increases the utility of distributed databases, creates efficiencies in data generation and use, and makes it possible to apply a broader spectrum of environmental information to any given problem. However, because of the many differences in data collection and management by different monitoring and measurement systems, creating interoperability among systems can be extremely difficult and time consuming. A key question is: how much standardization must there be among different systems for them to share information in a productive way? The issue of interoperability has been addressed by a number of coastal ocean observing systems, such as the Southeast Atlantic Coastal Ocean Observing System (SEACOOS) and its Information Management (IM) working group.

SEACOOS was initiated in September 2002 by a partnership of 12 institutions from North Carolina, South Carolina, Georgia, and Florida, as well as affiliations with over twenty additional federal and state agencies and institutions and the private sector (Seim et al., 2004). The founding institutions -- University of North Carolina at Chapel Hill (UNC-CH), Skidaway Institute for Oceanography (SkIO), University of Miami (UM), University of South Florida (USF), and University of South Carolina (USC) – had existing programs in place that produced data observations and/or model analyses of the coastal systems, and USC also was the headquarters

for the Centralized Data Management Office (CDMO) of the NOAA National Estuarine Research Reserve System (NERRS). These different programs served as a foundation for SEACOOS. The initial challenge was to bring these assets together and build on them in a way that advanced observations and understanding of elements of the coastal system, such as three-dimensional circulation, from the Outer Banks to the west Florida shelf.

As the SEACOOS IM working group set about establishing an IM system that promoted interoperability, it addressed the similarities and differences among its primary partners. It also monitored the emergence of a national system and worked to parallel developments in the Integrated Ocean Observing System (IOOS) (www.ioos.noaa.gov) Data Management and Communications (DMAC) (OceanUS, 2005). SEACOOS development focused on the processes and protocols needed to aggregate information from distributed sources, with an emphasis on capabilities to visualize and retrieve aggregated data from near real time data streams of *in situ* observations, model output, and remotely sensed imagery. The steps taken within SEACOOS to work with distributed monitoring systems could serve as a model for other sets of observations. The purpose of this paper is to briefly outline the most important steps and lessons learned in development of the SEACOOS IM system.

BACKGROUND AND RELATED WORK

Three fundamental principles served to guide system development: (1) The observation systems and associated databases would remain at the primary sources, with no attempt to centralize all distributed data in one database; (2) We would build upon existing resources and practices as much as possible to support participant autonomy, conserve resources, and promote rapid progress; and (3) We established an "open access" policy by ensuring that the data would be freely accessible in a timely manner, and IM developments would be constructed in a manner that could be readily adopted by other potential users.

The IM system was primarily composed of the data (real-time observations, remotely sensed data, and model output); a variety of functions that increased the information content of data (e.g. aggregation, analysis, map-based presentations, model input); and a variety of user-defined tools and applications. System integrity was assured through establishment of specific data standards, processes, and protocols required for quality assurance and quality control (QA/QC) as well as data and system documentation (IOC, 1993; NDBC, 2003).

Our first challenge was to develop the means to aggregate distributed data and demonstrate that they had been pulled together in a way that actually had utility. Accordingly we decided to generate a map-based product that displayed selected sets of aggregated data in a variety of forms. A visual map made it possible to assess relatively quickly whether data were represented accurately and whether the production protocols were working as intended. Map presentations also make good products for conveying information to a broader audience.

A central hub was established at USC, and more recently, additional nodes have been established at UNC-CH and USF to promote redundancy and system backup.

PROGRAM DESIGN

To establish interoperability, it was necessary to assess and decide upon a minimal set of standards to be used for data description, expression, and transport. This required a balance between retention of the variety of standards that were already in use by the data providers and the need to transition to a limited set of standards that could also be used by numerous ocean observation programs beyond SEACOOS. Through review and assessment of in-house practices and those of other environmental monitoring programs, database and transport standards were identified and documented. A common vocabulary, or "data dictionary," was established, which specified the naming and reference conventions, including observation position, time, and type of measurement (http://marinemetadata.org/references/seacoosdd).
Similarly, mechanisms for data transport were assessed, and for our purposes raw data files are generally transported via HTTP (Hypertext Transfer Protocol). Particularly useful for our map-based products are the Open Geospatial Consortium (OGC) (http://www.opengeospatial.org/) standards. These enable a set of web services (Web Map Services, WMS; Web Feature Services; WFS) that are compatible with most common GIS platforms. Illustrations of such applications, which include SEACOOS data, are at OpenIOOS (www.openioos.org) and the more recent OosTethys (www.oostethys.org).

The data file format selected was NetCDF, and specifications that were adopted for SEACOOS purposes are thoroughly documented in the SEACOOS NetCDF Common Data Language v2.0 (available at http://seacoos.org/documents/metadata). The convention addresses a wide range of data types, i.e. fixed point (e.g. mooring, offshore tower, and tide- and stream-gauge observations), 2D or 3D moving point (e.g. ship, surface drifter, or underwater autonomous vehicle measurements), fixed or moving profilers (e.g. bottom-mounted or ship-mounted acoustic profilers), and

remotely sensed observations (e.g. high-frequency radar, aircraft, satellites).

Some minimal metadata are included in the netCDF file format, but additional metadata are needed to fully understand the characteristics and quality of the data. SEACOOS contributed to an on-line browser-based tool, called MetaDoor, which assists in creation and publishing of XML metadata records, and it is available at http://www.carocoops.org/metadoor .

METHODS

A wide variety of processes, protocols, and software platforms have been assessed, and those that have been adopted as best practices are described in the SEACOOS Data Management and Visualization Cookbook (http://www.seacoos.org/documents/cookbook/). The Cookbook also provides information on implementation and troubleshooting.

.
CONCLUSIONS

SEACOOS contributed to interoperability through a range of developments in database structure, data standards, and data transport. The functional success of these developments is illustrated in two basic types of browser-accessible, map-based presentations. The first map type (Fig. 1) is a series of automatically generated maps that present the most recent data and are refreshed every hour (http://seacoos.org/Data%20Access%20and%20Mapping/cached-images). These merge a variety of distributed data to produce three primary products based on sea surface temperature (SST), wind, and water circulation. The second map type is interactive and generated by the user, who selects the specific data to be included (http://seacoos.org/Data%20Access%20and%20Mapping/InteractiveMap/). A page has also been developed that displays data animations, showing changes in variable measurements with time. This allows the user to select GIS layers, scale, observation sites to graph, and time steps. Developers that wish to implement this technology in their systems can access animation templates (http://caro-coops.org/bb/viewtopic.php?t=326&highlight=), as well as instructions for how to construct individually tailored animations.

DISCUSSION AND RECOMMENDATIONS

A fundamental goal of SEACOOS is IM interoperability. This was achieved with the primary partners by establishing a set of standards that enabled data transport and aggregation, as well as production of map-based products from aggregated data. National programs working towards standards-based ocean-data interoperability include OpenIOOS, OOSTEthys, and the IOOS Observations Registry. Progress is being made in finding paths to interoperability with federal data providers, such as the NOAA National Data Buoy Center (NDBC). There is tremendous potential value in achieving interoperability with state agencies, as many have valuable, extensive sets of databases, such as those related to water quality. Because of the enormous potential for interoperability to enhance our abilities to address increasingly urgent environmental issues, such as those related to water issues, we should make a collective effort to address the standards issue within the academic, state, federal, and private sectors.

Progress towards interoperability within the coastal ocean IM community has resulted in some significant "lessons learned" that may aid collaborative efforts addressing water resource issues. First, data managers and programmers from various institutions are likely to be able to form a highly productive, networked community that thrives in shared problem-solving. Efforts should be made from the outset to bring these technical personnel together and establish a dynamic, collaborative, virtual community. Second, Information Management should be recognized as a core function and be supported accordingly. Interoperability cannot develop on "leftovers." Third, standards must be identified, and that takes committed effort and consensus. Fourth, if consistent and reliable information is a priority, appropriate redundancy and back-up must be established. Finally, real-time data provide information critical for making sound decisions about near-term events; however, historical databases are essential for understanding changing environments and the nature of human impacts. Both types of data should be accommodated in regional IM infrastructures.

ACKNOWLEDGMENTS

LITERATURE CITED

International Oceanographic Commission._1993. IOC Manuals and Guides No. 26 -- Manual of Quality Control Procedures for Validation of Oceanographic Data., UNESCO, http://www.iode.org/index.php?docID=874&option=com_oe&task=viewDocumentRecord .

National Data Buoy Center,2003. Handbook of Automated Data Quality Control Checks and Procedures of the National Data Buoy Center, NDBC Technical Document 03-02, 2003, http://www.ndbc.noaa.gov/handbook.pdf , NOAA.

Seim, H., B. Bacon, C. Barans, M. Fletcher, K. Gates, R. Jahnke, E. Kearns, R. Lea, M. Luther, C. Mooers, J. Nelson, D. Porter, L. Shay, M. Spranger, J. Thigpen, R. Weisberg and F. Werner. 2004. SEACOOS – A model for a multi-state, multi institutional regional observation system. Mar. Technol. Soc. J., 37: 92-101.

OceanUS. 2005. Data Management and Communications Plan for Research and Operational Integrated Ocean Observing Systems. (Accessed 5/28/2008 at http://dmac.ocean.us/dacsc/imp_plan.jsp;jsessionid=2030311176538409441).
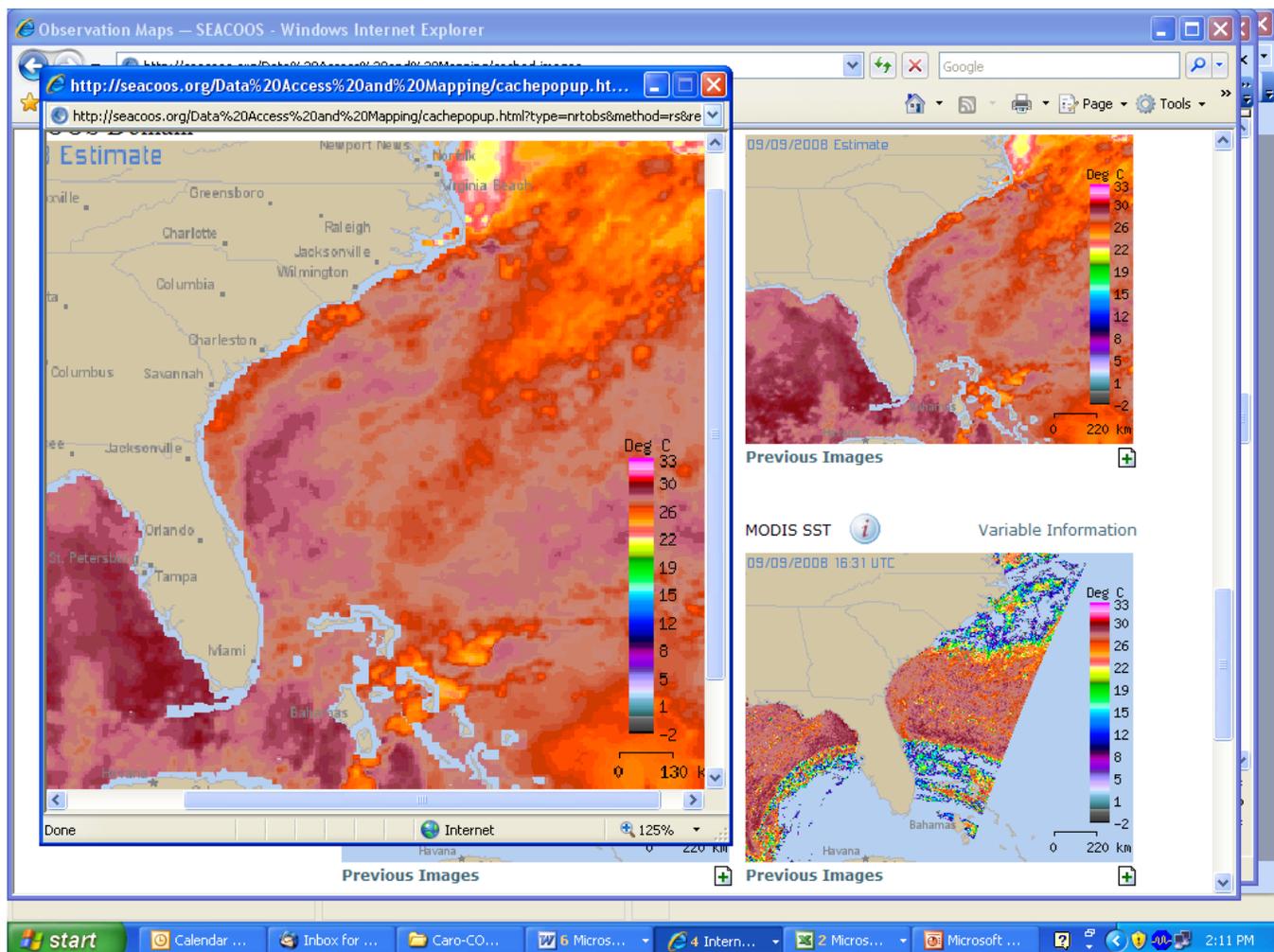
Figure 1: Interpolated satellite imagery of sea surface temperature captured on September 9, 2008.