

Integrating heterogeneous databases in Ocean Biogeographic Information System

Yunqing Zhang and J. Frederick Grassle

Institute of Marine and Coastal Sciences, Rutgers, the State University of New Jersey
71 Dudley Road, New Brunswick, NJ 08901, USA

E-mail: phoebe@imcs.rutgers.edu

Ocean Biogeographic Information System (OBIS) is an on-line, open-access, globally-distributed network of systematic, ecological, and environmental information systems. Collectively, these systems operate as a dynamic, global digital atlas to communicate biological information about the ocean and serve as a platform for further study of biogeographic relationships in the marine environment. Emphasis is on accurately-identified, species-level, geo-referenced abundance data. Heterogeneity exists at all levels of data management of OBIS members, ranging from the underlying computer hardware and operating system to data format and semantic modeling. Smooth interoperation of these heterogeneous data systems is critical to OBIS development. The interoperability is accomplished after a distributed infrastructure and interoperation standards are established. In OBIS structure development, globally distributed data nodes are established for experts to store, manage, version and quality-control data in their specialty fields. Meanwhile, a portal server is placed between data nodes and the end user to support one-stop data “shopping” via a uniform user interface. Taxonomy name service, environmental mapping service, genetic information service, biogeographic modeling service, and education programs are also being developed, tapping into system-wide data resources. Such a structure combines the effectiveness and scalability of distributed systems with the efficient user access offered by an information portal. To facilitate communication within the distributed system, the OBIS technical working group selected HTTP as the communication protocol and XML as the data exchange format. XML technology is enormously scalable as long as data content is not contracted. HTTP and XML technology have broad industrial support, and system development based on them is less time-consuming and technically demanding than competing approaches. A call interface over HTTP and a global federation schema based on the Open GIS standard (GML) are also defined. OBIS node servers and the portal server implemented this set of interfaces and standards so that the OBIS portal can provide a Web-based transparent search capability across all databases. The successful operation of the interoperable OBIS demonstrates that XML can be used to facilitate not only the exchange of structured data, but also semantic data integration through the construction of a global schema. The rapid development of OBIS calls for extension of its global schema to support complex spatial data types and operations. Incorporating both grid and vector based data into one semantically-integrated system is one major problem that the oceanography and GIS community have to solve together. Whether a pure XML-based system or a mixed system should be adopted to deliver the spatial operations is a research topic that needs to be addressed in the near future.