Toward Improved Data Availability: The Biological & Chemical Oceanography Data Management Office

BCO-DMO

Cynthia L. Chandler[1], Robert C. Groman[2], M. Dicky Allison[2], David M. Glover[1], and Peter H. Wiebe[2]

[1]Marine Chemistry and Geochemistry Department, [2] Biology Department, Woods Hole Oceanographic Institution, Woods Hole MA 02543

Abstract

Data availability speeds knowledge discovery, increases the efficiency of the scientific enterprise, and rewards those that support its collection through better decision support at all levels. The first step in improving data availability is to increase the volume of data published electronically. But achieving this step alone does not guarantee the data will be usable. To make data truly available for scientific use, they must be discoverable and discernible. Whether the driver of a data access client is machine or human, availability of metadata is one of the most important agents of success.

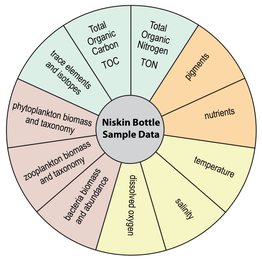
The Biological and Chemical Oceanography Data Management Office (BCO-DMO) was created to serve investigators funded by the NSF Biological and Chemical Oceanography Sections and to support the scientific research community through improved access to ocean science data. The BCO-DMO is a combination of the formerly independent data management offices for U.S. GLOBEC and U.S. JGOFS programs. One of the main goals of the BCO-DMO is to develop an information management system from which marine biogeochemical and ecological data and information developed in the course of scientific research can easily be disseminated and data can be protected and stored for short and intermediate time intervals.

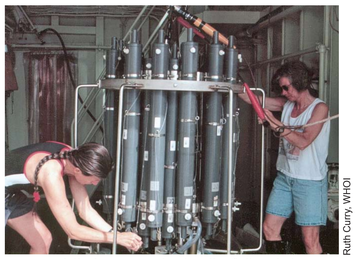
Another of the goals of the BCO-DMO is to facilitate regional, national, and international data and information exchange. Biogeochemistry and ecosystem research projects are inherently interdisciplinary and benefit from improved access to well-documented data. The BCO-DMO is developing an open access data system that uses public domain software: project and data set metadata records designed to support accurate re-use of the data are stored in a relational database (MySQL); data are stored in or made accessible by the JGOFS/GLOBEC object-oriented, relational, data management system; and access to geospatial data will be provided via any standard Web browser client through a GIS application (Open Source, OGC-implementing MapServer). Storing metadata in a relational database permits data set descriptions to be generated in compliance with a variety of metadata content standards and provides mechanisms for exchange of data with other data repositories.

Reconstituting the Ocean: Best Practices and Data Integration

The main objective for BCO-DMO staff members is to support the scientific community through improved access to ocean science data. Data integration, the process of combining data from disparate data sets, requires access to data and metadata. A sampling device such as a CTD/Rosette package (below) or the MOCNESS (far right) is capable of generating a variety of data sets of interest to researchers. BCO-DMO staff members work closely with investigators from the time of initial funding through data publication—'proposal to preservation' — to encourage data management best practices. For example, results from a typical research cruise should include the items listed below.

- a project description and science plan
- a post-cruise report
- a sampling event log
- an inventory of data sets acquired
- quality-controlled data and metadata



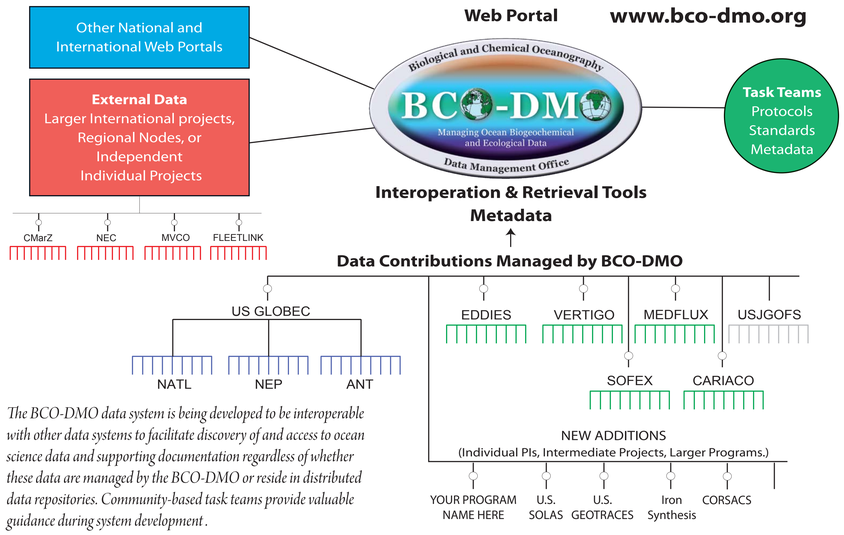


Ruth Curry, WHOI

A rosette full of Data. *Cyndy Chandler (left) and Theresa McKee draw oxygen, salinity and nutrient samples from Niskin bottles after a CTD station has been completed during a May 2000* Knorr *cruise.*

For a multi-disciplinary sampling program conducted aboard a research vessel, a sampling event log and cruise report constitute an important part of the final results and greatly enhance the value of the published data. Documented sampling and analytical protocols, instrument descriptions, quality control and post-processing procedures complete the metadata records that will enable subsequent data integration.

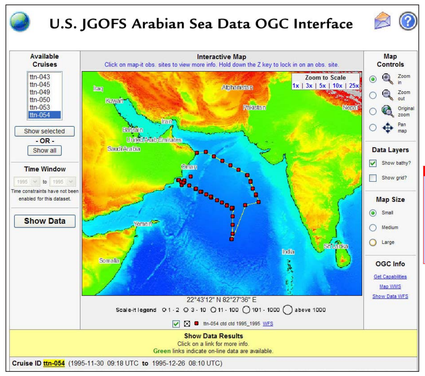
Data Availability

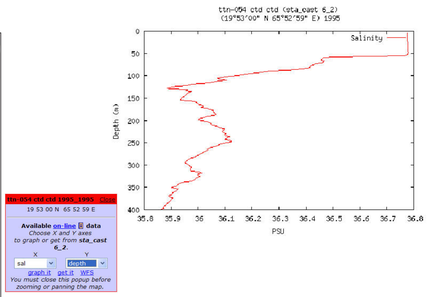


The BCO-DMO data system is being developed to be interoperable with other data systems to facilitate discovery of and access to ocean science data and supporting documentation regardless of whether these data are managed by the BCO-DMO or reside in distributed data repositories. Community-based task teams provide valuable guidance during system development.

Interoperability and Visualization

One important design specification for the new BCO-DMO data system is that components should be chosen to increase interoperability with other data systems regardless of boundaries associated with disparate research disciplines, funding agencies, or nationalities. The current data management system, using the JGOFS/GLOBEC distributed, object-oriented, relational data system, is being enhanced to provide a more uniform, project-independent approach to storing metadata and data. The user interface to the previous data system was based predominantly on tabular display of selected data and human readable metadata. Several different types of user interfaces are being developed for the new system. Many of the data sets published by the BCO-DMO can be represented in a geospatial context (e.g. mapped according to latitude and longitude), and developing a single geospatial interface to the full BCO data collection is an important priority. We hope to improve overall system interoperability by using well-defined term names from controlled vocabularies, dictionaries and ontologies, and through support for metadata standards such as those developed by the Federal Geographic Data Committee (FGDC), the American National Standards Institute (ANSI) and the International Standardization Organization (ISO). Data access options include support for community-wide standards such as the Open-source Project for a Network Data Access Protocol (OPeNDAP) framework (that provides transparent access to oceanographic data) and Open Geospatial Consortium (OGC) compliant Web Services provided by the OGC-implementing MapServer interface for geospatial data. Data and supporting documentation are available to research investigators and the public and the system enhancements provide additional visualization capabilities and help to ensure that the new BCO-DMO data system will be interoperable with data exchange systems at other repositories.

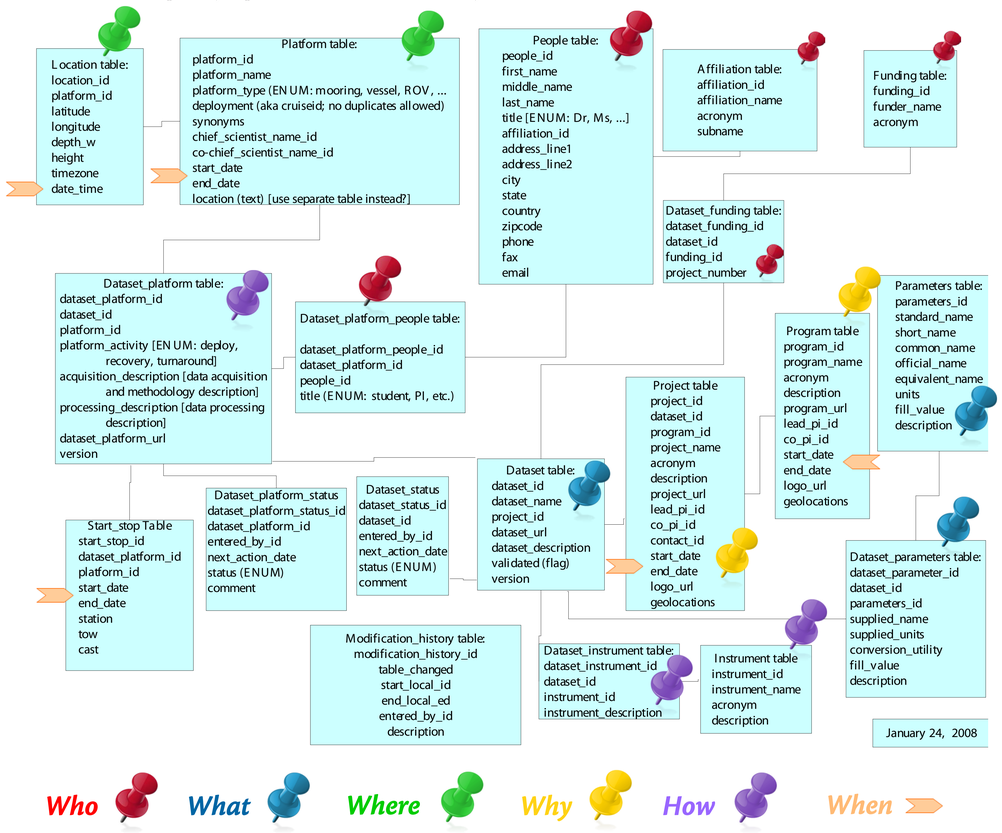




The MapServer (University of Minnesota, mapserver.gis.umn.edu) interface facilitates data discovery and provides data access via the Open Geospatial Consortium (OGC) Web Mapping Service (WMS) and Web Feature Service (WFS) protocols.

Metadata Database Schema

Robust metadata records are the key to accurate use of data. Metadata records capture the information required to answer the who, what, where, why, how and when questions that are asked about a data set. It is important to know who collected, analyzed and contributed the data and where, when and how those data were acquired and subsequently analyzed and processed. Ideally, term dictionaries and controlled vocabularies are used to populate the metadata database fields. Access to metadata and supporting documentation aids data discovery (through development of powerful search engines) and helps to ensure efficient and accurate use of data. Data reuse is especially dependent on metadata availability.



Conclusions

Peter Wiebe, WHOI

Hauling in the Data. *Scientists haul in the 10 meter MOCNESS nets after a deep tow during an April 2006 cruise aboard the Ron Brown. Data sets from a MOCNESS sampling system may include zooplankton species composition, abundance, temperature, salinity, depth, and possibly oxygen, fluorescence, transmissometry and downwelling light.*

Decades of experience managing data for multi-disciplinary ocean research projects has informed our current approach to data and information management in support of marine biogeochemical and ecological research. Knowledge gained from valuable lessons learned is reflected in the new BCO-DMO system – a system designed to be simple, reliable, flexible and modular. The publicly available BCO-DMO data collection represents one mechanism enabling investigators to share and exchange scientific research results. The ability to exchange data fosters interdisciplinary interactions within a variety of research areas including marine biogeochemistry, ecological modeling and forecasting, global carbon budget analyses, climate change research, and ecosystem and biodiversity studies.

Presented at 2008 NASA Carbon Cycle and Ecosystems Joint Science Workshop (CC&E JSW), April 28-May 2, 2008
University of Maryland Inn and Conference Center
3501 University Boulevard East, Adelphi, MD 20783