BCO-DMO

Biological & Chemical Oceanography Data Management Office

ESSENTIAL PARTNERSHIPS IN THE DATA MANAGEMENT LIFE CYCLE

IN21C - 1705
Fall AGU 2015

Danie Kinkade, Molly D. Allison, Cynthia L. Chandler, Nancy J. Copley, Steve Gegg, Robert C. Groman, Shannon Rauch

Woods Hole Oceanographic Institution (WHOI), Woods Hole, MA USA

Value to Research Community

The Research Investigator Partnership

Today's geoscience research efforts can rapidly produce an unprecedented volume of multidisciplinary data that can pose management challenges for the facility charged with curating that information. How do these facilities achieve efficient data management in a high volume, heterogeneous data world? Partnerships are critical, especially for small to mid-sized data management offices, such as those dedicated to academic domain research communities. The idea of partnerships can encompass a wide range of collaborative relationships aimed at helping these facilities meet the evolving needs of their communities. However, one basic and often overlooked partnership in the data management process is that of the information manager and the Principal Investigator (PI) or data originator. Such relationships are critical in discerning the best possible management strategy, and in obtaining the most robust metadata necessary for reuse and interoperability of multidisciplinary datasets. Partnerships established early in the data life cycle enable efficient management and dissemination of data in high volume and heterogeneous formats.

The Biological and Chemical Oceanography Data Management Office (BCO-DMO) was created to fulfill the data management needs of PIs funded by the NSF Ocean Sciences Biological and Chemical Sections, and Division of Polar Programs. Since its inception, the Office has relied upon the close relationships it cultivates with project PIs in order to provide effective data management for a wide variety of ecological and biogeochemical oceanographic data. Here we highlight some of the successful partnerships BCO-DMO has made with individual and collaborative investigators within the GEOTRACES community. Partnerships such as these are an essential component in the effective management of individual investigator data. Through these and other strategic partnerships, domain specific repositories such as BCO-DMO create 'smarter data' to support the research community's needs.

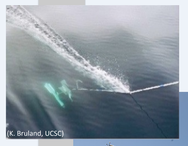
GEOTRACES

GEOTRACES is an international program which aims to improve the understanding of biogeochemical cycles and large-scale distribution of trace elements and their isotopes in the marine environment. Scientists from approximately 35 nations have been involved in the program, which is designed to study all major ocean basins over the next decade. (www.geotraces.org)

(K. Bruland, 2008)

U.S. GEOTRACES contributes to the mission of the GEOTRACES program to identify processes and quantify fluxes that control the distributions of the key trace elements and isotopes in the ocean, and to establish the sensitivity of these distributions to changing environmental conditions.

A variety of data are collected from shipboard and deployed instruments, and laboratory analyses.



(K. Bruland, UCSC)

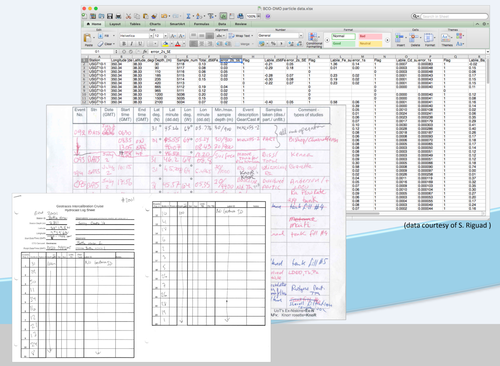
(J. Fitzsimmons, MIT)

(ct.rsmas.miami.edu)

(J. Swift, UCSD)

(K. Kostel, WHOI)

Files and metadata are submitted to BCO-DMO in various states of completeness and usability.



(data courtesy of S. Riguad.)

Other Strategic Partnerships

Recognizing this model does not always scale, and does not stop at the PI-data manager relationship, additional strategies and partnerships are necessary to create 'smart data'. Technologies adopted in close collaboration with groups that support other research communities help to integrate data repositories to build a valuable knowledge network. BCO-DMO has evolved to include Semantic Web technologies and Linked Open Data to connect our data with complementary information in other relevant repositories; the results of which enhance data discoverability, and access, and quality. 'Smart data' become a more interoperable and valued community resource.

Rolling Deck to Repository (R2R)

rvdata.us

BCO-DMO and R2R use Linked Data to exchange resources curated at each of the repositories.

NATURAL ENVIRONMENT RESEARCH COUNCIL

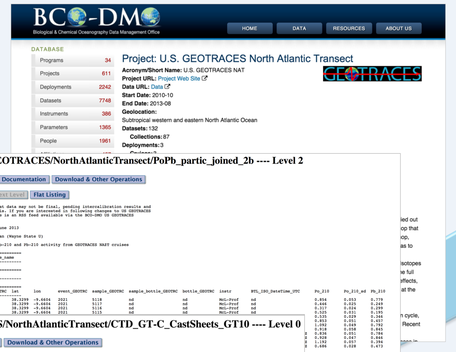
BCO-DMO and R2R federate controlled vocabulary terms (e.g., Instruments) using the NERC Vocabulary Server (NVS 2.0; Leadbetter, 2013).

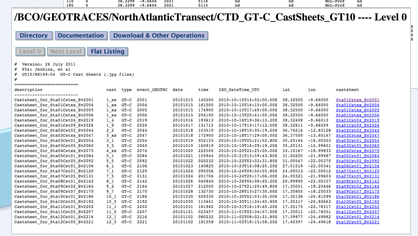
NCEI

NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION

BCO-DMO partners with the National Center for Environmental Information (NCEI) for long-term preservation of curated BCO-DMO data.

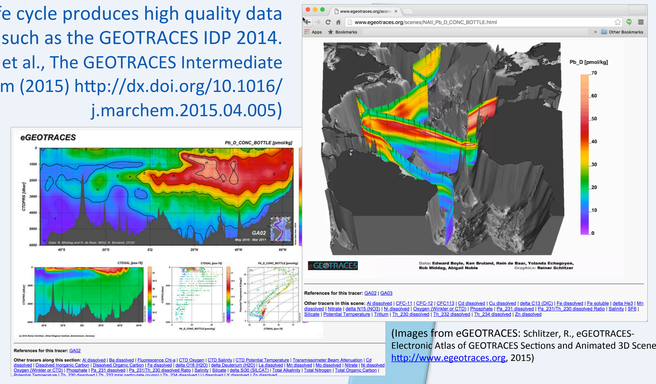
Documentation assembled for projects and cruises provides context for complex data.





Data managers work closely with data originators to create robust metadata, perform gross quality control, and determine the best way to display their data.

Value added along the data life cycle produces high quality data available for use in products such as the GEOTRACES IDP 2014. (Data available from: Mawji, E., et al., The GEOTRACES Intermediate Data Product 2014, Mar. Chem (2015) http://dx.doi.org/10.1016/j.marchem.2015.04.005)

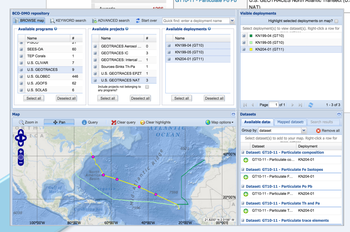


(Images from eGEOTRACES: Schlitzer, R., eGEOTRACES-Electronic Atlas of GEOTRACES Sections and Animated 3D Scenes, http://www.egeotraces.org, 2015)

BODC

GEOTRACES International Data Assembly Centre

Curated BCO-DMO 'smart data' is contributed to the International GEOTRACES data office at the British Oceanographic Data Center (BODC). These data are available with those from other international GEOTRACES investigators.



BCO-DMO uses Semantic Web technologies to enhance search interfaces for improved data discoverability and access.

ACKNOWLEDGEMENTS

This work is supported by NSF grant: OCE-1435578. We acknowledge the investigators who contribute data to BCO-DMO and the data managers who work to make those data available. Julie Allen and Katherine Joyce (WHOI) helped develop the BCO-DMO data system user interfaces and Charlton Galvarino (Second Creek Consulting, LLC) helped develop the map interface.

Curation Level

Strategic curation along the data's lifecycle increases the usability of individual datasets to a level where interoperability and integration are achievable

bco-dmo.org