BCO-DMO Progress in Data Discovery and Access: The Next Steps

Biological & Chemical Oceanography Data Management Office

Poster: 1707
Session: 071
Ocean Sciences Meeting
February 24-28, 2014

Robert C. Groman, Molly D. Allison, Cynthia L. Chandler, Nancy J. Copley, Stephen R. Gegg, David M. Glover, Danie B. Kinkade, Shannon M. Rauch, Adam Shepherd and Peter H. Wiebe
Biological and Chemical Oceanography Data Management Office, Woods Hole Oceanographic Institution (WHOI), Woods Hole, MA USA

Abstract

Data are needed for analysis, synthesis, initial conditions and verifying model output, and advancing new theories on how the world works. The data need to be of high quality, collected and processed by established and community accepted standards, discoverable, and accessible. There has been progress recently towards making data discoverable and accessible because of a combination of government requirements and higher expectations, but there is still much to do. The Biological and Chemical Oceanography Data Management Office (BCO-DMO) works in partnership with ocean science investigators to serve data from research projects funded by the Biological and Chemical Oceanography Sections and the Division of Polar Programs Antarctic Organisms & Ecosystems Program at the U.S. National Science Foundation. While BCO-DMO's data management model addresses many of the obstacles challenging the goal of data interoperability, there is still much to accomplish. This presentation targets where we are going and how we plan to get there.

Dialogue: April 2, 20xx

You: Show me the data that compares favorably with my MANTL version 7 model run from last night.

Computer: Can't because data aren't federated. Where should I look for the data? Anyone's data?

You: No. Only use data from trusted, authoritative sources.

Computer: But are temp1, air temperature and mytemp data comparable to your sea surface temperature?

You: Not sure. Check their ontologies.

Computer: Okay, but Smith's data comes from a new instrument with differences in acquisition and processing steps. Should I use these data along with all the others?

You: Why is this still so complicated?

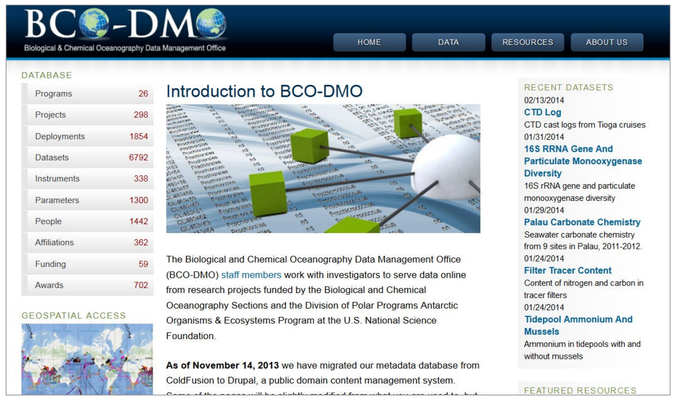


Figure 1: Current BCO-DMO home page providing text-based access to the metadata and data managed by BCO-DMO.

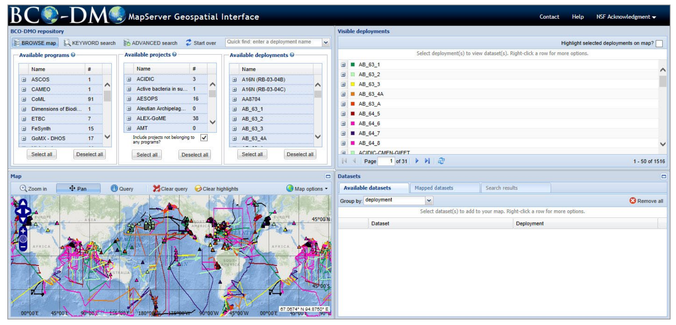


Figure 2: BCO-DMO's geospatial access is based on MapServer originally developed at the University of Minnesota.

Support investigators

- More support in preparation of Data Management Plan
- Continue one to one interactions for the entire data life cycle, from proposal to collection to analysis to publication and data use and reuse (Higgins, 2012; Michener et al., 2011; Strasser et al., 2011)
- Uploading data (Figure 3)
- More data quality checking –range and bounds checking
- Help with dark data – long tail of data
- More widespread use of Digital Object Identifier (DOI)
- Data citation credit

Improve data searching

- Use of ontologies to help find data using common or similar terminology
- Linked Data – start of federated data
- Interoperability of distributed systems – combining Linked Data and Semantic Web technology
- Geospatial searching supported by faceted search features (Figure 4)

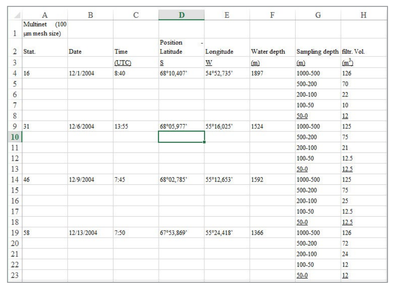
Assess fitness for purpose

- New instruments necessitate new ways to look at data
- New data visualizations will be needed (Figure 5)

Access data and metadata

- Add Open-source Project for a Network Data Access Protocol (OPeNDAP)
- Support International Standards Organization's ISO19115-2 metadata standard
- Download multiple datasets:ASCII tab or comma separated; MATLAB; Network Common Data Form (netCDF); Ocean Data View (ODV), and Keyhole Markup Language (KML)

Figure 3a-b: Originally contributed spreadsheets can take on an infinite number of organizations and column naming conventions (figure a). In this case, the data columns are labeled clearly, but the data should be re-organized (figure b) before serving.



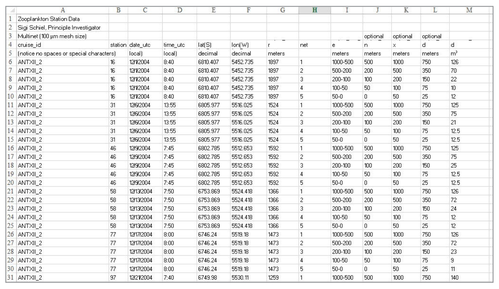


Figure 4: The beta version of the BCO-DMO faceted search option allows selection criteria to be chosen in any order.

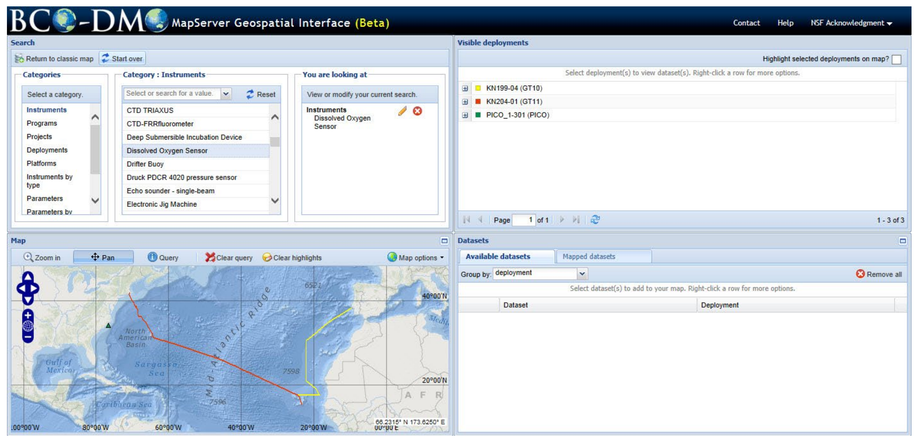
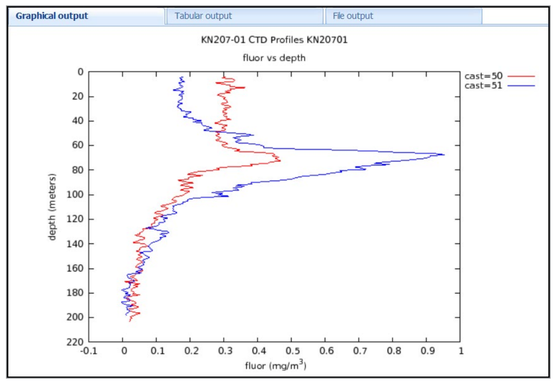
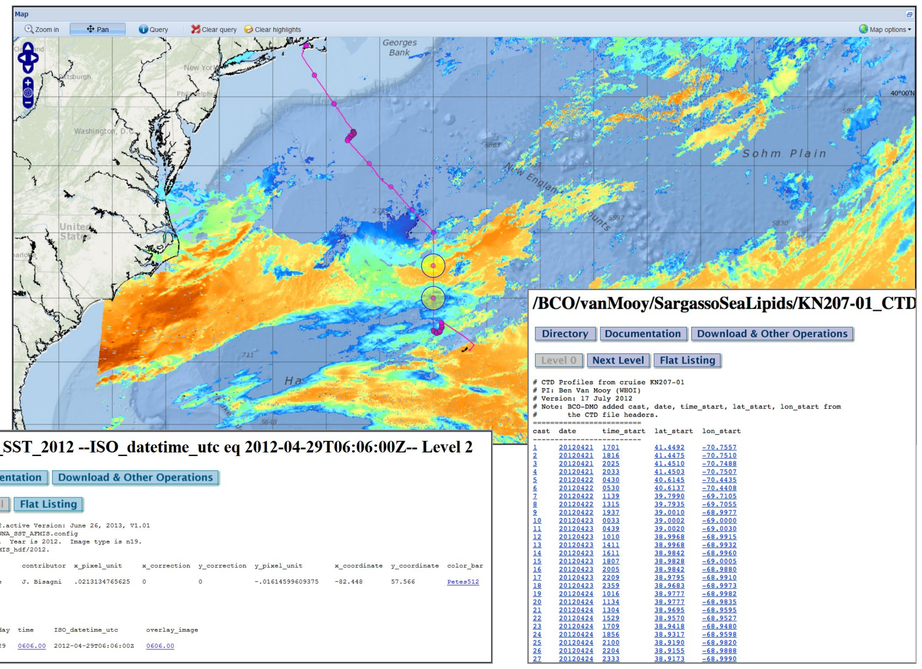


Figure 5a-d: Combining data from multiple datasets: Sea surface temperature from a satellite image are overlain with fluorometry data from two stations in the North Atlantic. Data courtesy of J. Bisagni and B. vanMooy





References

Higgins, S. (2012). The lifecycle of data management. In Pryor, G. (ed.), Managing research data (pp. 17-46). London: Facet Publishing.

Michener, W. K., Porter, J., Servilla, M., & Vanderbilt, K. (2011). Long term ecological research and information management. Ecological Informatics, 6(1), 13-24. doi: http://dx.doi.org/10.1016/j.ecoinf.2010.11.005

Strasser, C., Cook, R., Michener, W., Budden, A., & Koskela, R. (2011). DataONE promoting data stewardship through best practices. In Jones, M.B. & Gries, C. (eds.), Proceedings of the Environmental Information Management Conference 2011 (EIM 2011) (pp. 126-131). University of California. Retrieved from https://eim.ecoinformatics.org/eim2011/eim-proceedings-2011/view

Acknowledgments

BCO-DMO is funded by the National Science Foundation. We acknowledge the work done by all of the investigators who contribute their data to BCO-DMO. The user interfaces to the BCO-DMO data system were developed in collaboration with Julie Allen and Katherine Joyce (WHOI). The geospatial interface to the BCO-DMO data system was developed in collaboration with Charlton Galvarino (Second Creek Consulting, LLC).

www.bco-dmo.org
info@bco-dmo.org