

Coping with Your Data: WHOI Data Solutions and Resources Workshop

July 26, 2013

Workshop Summary

The “Coping with Your Data” workshop brought together scientists, technical staff, students, and program managers to discuss challenges and opportunities in accessing, handling, and sharing data. The approximately 40 participants were mainly from WHOI but also included several participants from the local science community (NOAA, USGS). The workshop was convened by WHOI’s Ocean Informatics “Phase 2” Advisory Group: Stace Beaulieu (Biology), Cyndy Chandler (MCG), Joe Futrelle (AOPE), and Lisa Raymond (Co-Director MBLWHOI Library). The workshop consisted of several brief introductory talks, several “lightning” talks highlighting informatics projects at WHOI, demonstration tables, and a lunchtime discussion. The agenda, list of participants, list of demonstration tables, and notes contributed during the workshop are posted in a Google document (https://docs.google.com/document/d/1-R8e0OKqzM1mJQDTNu_7uqwTmV-Dfdl0OuCC7N4aHqc). The workshop had several immediate outcomes including specific questions answered (e.g., contributing data to CLIVAR and Carbon Hydrographic Data Office) and connections made between people with different expertise (e.g., scientist and technologist for high volume image processing).

The workshop opened with an introduction by Lisa Raymond and a report by Andy Maffei (CIS) on WHOI’s Ocean Informatics “Phase 1” Working Group, active from 2009-2013. This Working Group was sponsored by WHOI’s Director of Research to recommend a strategy and an action plan for a robust data informatics infrastructure at WHOI. A full report, documenting the projects and outcomes generated in Phase 1 is posted online (<http://www.whoi.edu/DoR/special-projects/ocean-informatics-working-group>).

Next, Dave Garrison, manager of the Biological Oceanography, Ocean Acidification, and LTER programs at the National Science Foundation, and Jesse Ausubel, science advisor to the Sloan Foundation, gave brief presentations on current status and future needs for data management in their respective agencies. Garrison highlighted the Biological and Chemical Oceanography Data Management Office (BCO-DMO), hosted at WHOI. In terms of the challenges of “Big Data,” Ausubel offered the concept of macroscopes – new tools to help us share data on a large scale and help us see patterns/ gain understanding from data that are very large in scope.

Beaulieu concluded the opening talks by introducing the WHOI Ocean Informatics “Phase 2” Advisory Group and goals for 2014: 1) Strengthen connections to expertise within WHOI and our local science community, 2) Become aware of broader community efforts and external opportunities that may lead to new funding, 3) More student involvement, and 4) Help with data and information management plans.

Then, five lightning talks were delivered in rapid succession, followed by a question and answer period. These 5-min talks highlighted informatics in the Imaging Flow Cytobot (IFCB) dashboard, Tetherless

World Constellation (TWC) at Rensselaer Polytechnic Institute (RPI) use case methodology, reproducibility for NOAA ecosystem status reports, smart-search capability for the BCO-DMO database, and HabCam image annotation. Much of the discussion following the lightning talks centered on the importance of the up-front work in building a use case and concept maps for successful informatics partnerships and technology development.

Participants then were free to browse and interact with the following tables with software demonstrations:

- HabCam, Burton Shank (NOAA)
- IFCB, Joe Futrelle and Heidi Sosik
- Ecosystem Status Report and IPython Notebooks, Stace Beaulieu
- Publishing your data with Digital Object Identifiers (DOIs), Ann Devenish and Lisa Raymond
- EarthCube and the MVCO, Janet Fredericks
- BCO-DMO, Cyndy Chandler and Danie Kinkade
- R2R Eventlogger, Cindy Sellers

Lunch table discussions were focused on the following questions (derived from responses to our overarching question posed prior to the workshop: “What are the biggest barriers for you to work with your data?”):

What file formats should I use to manage my data?

Where can I get funding for data management, migration, and preservation?

How should I migrate/preserve my data?

What metadata standards/controlled vocabularies should I use?

What tools are available to organize and make my data accessible?

Should I embargo my data or make it open access?

How do I cite data and assure proper attribution?

Who or how can I get help to cope with my data?

The discussions ranged from specific needs, such as funding for migration of data that are on media that is degrading, to the broad concepts of open access and informatics training. There was a recommendation to involve informatics expertise when writing proposals, so that informatics is integrated before a project starts. A question was posed: “Should WHOI have an institute for ocean informatics?” The Advisory Group will follow up with each individual who submitted an email or index card for “My Data Challenge” and, over this next year, will evaluate further these questions discussed above.