UNOLS Data Best Practices

October 3, 2008

Committee Members

Arko, Robert	Lamont-Doherty Earth Observatory
Case, James	University of New Hampshire
Chandler, Cyndy	Woods Hole Oceanographic Institution
Chapman, Piers	Texas A & M University
Fanning, Bill	University of Rhode Island
Forcucci, David	U.S. Coast Guard
Graybeal, John	Monterey Bay Aquarium Research Institute
Johnson, H. Paul	University of Washington
Measures, Chris	University of Hawaii
Mendelssohn, Roy	NOAA/NMFS/SFWSC
Miller, Stephen	Scripps Institution of Oceanography
Smith, Shawn	Florida State University

Committee was established in July 2007

- Report to the Council on current community practices for the collection of data and metadata at sea, identify best practices, and make recommendations for improvements.
- Since their last report on October 11, 2007, the Committee has met twice in person, at the AGU Fall Meeting (December 11, 2007) and the AGU Ocean Sciences Meeting (May 5, 2008), as well as corresponded by phone and email.

Findings

- An initial survey of committee members quickly revealed that no standard community practice for reporting data/documentation from ship to shore currently exists. Where such practices do exist, they are typically driven by –
  - 1. A particular operating institution with a “tradition” of routinely documenting and archiving data from its ship(s), or
  - 2. A specific program mandate (e.g. Ridge 2000 or MARGINS) that requires individual investigators to document and archive their data.

Conclusions

- Given that no standard community practice exists, the committee's primary work over the past year has been to develop recommendations for such practices going forward. Our work has been guided by these broad principles –
  - 1. The community is best served if information is *routinely reported by every platform to a central repository in a standard format; and*
  - *2. It's easier to train a dozen operators than a thousand scientists.*

Recommendations

- The committee recommends this initial set of data/documentation be routinely reported for every leg:
  - 1. Cruise summary (ship name, cruise id, dates/ports, personnel, data inventory, etc);
  - 2. Navigation (ship track i.e. time & position);
  - 3. Cruise event log (“everything over the side”, including both science and engineering).

How to implement Recommendations

- Much of this documentation is already *routinely collected by the operators and/or the UNOLS Office, and needs only to be standardized and centralized.*
- *Every ship in the fleet today uses GPS and could report navigation data to a central repository.*

Cruise Level Metadata

- The committee spent considerable time discussing and revising a proposed minimum standard for summary ("cruise-level") metadata. The current draft schema (version 1.6a) is attached. Our work was guided by the following design constraints:
  - 1. Every cruise must be assigned a unique and persistent identifier, including transits that are charged days. This identifier becomes the primary key to related documentation, data, and publications;
  - 2. A significant number of metadata elements can be pre-populated by the UNOLS Office or other shoreside repository, and/or pre-populated from a "vessel profile" that maintains a chronology of the standard instruments installed on each ship;
  - 3. Metadata elements should use controlled vocabularies wherever possible, ideally drawn from existing community standards, and synchronized with Ship Time Request and Scheduling System (STRS) vocabularies;
  - 4. The schema should be sufficiently generic to extend beyond UNOLS to potentially include other NSF-funded research vessels (USAP, USCG, etc); 5. *Navigation is metadata. Barring exceptional circumstances such as classified missions, navigation should be made routinely and immediately available.*

Why transmit navigation data ashore?

- Once ships have reported their navigation to a central repository, further shoreside processing could be done such as:
  - 1. reformatting, resampling, editing, annotating;
  - 2. calculating control points (abstracted trackline) and bounding boxes;
  - 3. evaluating quality and comparing systems.

The Way Forward

- The committee reported results and solicited feedback at the 2007 RVTEC meeting in Monterey, and will do so again at the 2008 RVTEC meeting in Tallahassee.
- Committee members are also involved with several other closely-related data management activities, including the –
  - NSF Legacy of Ocean Exploration (LOE) project
  - NSF Rolling Deck to Repository (R2R) project.

NSF Legacy of Ocean Exploration (LOE) project

- The LOE project, funded by a five-year NSF award (2005-2010), held a community meeting in Palisades, NY, on September 3-5, 2008, to address data management in the academic fleet.
- A significant number of UNOLS managers and technicians from a broad cross-section of operators attended, as well as the NSF OITS program director and the UNOLS executive secretary.
- The meeting's goals were –
  - 1. Review data management current practices on NSF-supported research vessels;
  - 2. Develop and discuss options for routine, standard production of cruise documentation and underway data, and delivery to a shoreside repository in near-realtime;
  - 3. Plan next steps and candidates for pilot projects to advance community standards.
- The meeting produced significant consensus, and a detailed report is forthcoming.
- Meeting materials and results are posted at *http://mgds.ldeo.columbia.edu/projects/legacy/*

NSF Rolling Deck to Repository (R2R) project.

- The R2R project, funded by a 1-year NSF supplement (2008-2009), is a collaborative pilot project between LDEO, SIO, and WHOI to develop a prototype “data discovery system” for NSF-supported research vessels.
- The project’s goals are:
  - 1. Establish the infrastructure for a central shore side repository, and
  - 2. Work with ship operators to develop and test procedures for the routine delivery of standard documentation and data products to the repository.
- The R2R project will build upon the UNOLS Data Committee’s recommendations for standard cruise documentation and data products, as well as results from the LOE meeting.
- Further community discussion is planned for the December 2008 AGU meeting in San Francisco, in association with posters and presentations in Sessions IN12 and IN19.