Research Vessel Data Management Progress Report

Robert Arko

RVTEC Annual Meeting
October 28, 2008

Outline:

1. UNOLS Data Management Committee update.
2. Legacy of Ocean Exploration (LOE) meeting.
3. Rolling Deck to Repository (R2R) pilot project.

1. UNOLS Data Management Committee update.

Committee Charge: "*Report to the Council on current community-wide best practices in data and metadata capture when collecting data at sea, and make recommendations for improvements.*"

Committee focus on initial/core set of documentation to be routinely reported:

1. Cruise summary (ship name, cruise id, dates/ports, personnel, data inventory, etc.)
2. Ship track (i.e. GPS is enough)
3. Cruise event log (both science and engineering)

"Cruise summary" metadata schema *(current v1.6a)* – guiding principles –

1. Every leg should be assigned a unique and persistent identifier;
2. Instrument metadata should be maintained as a separate "Vessel profile";
3. Controlled vocabularies should re-use community standards, and align with STRS + POGO;
4. Schema should be sufficiently generic to extend beyond UNOLS;
5. Navigation is metadata, and should be immediately available.

2. Legacy of Ocean Exploration (LOE) meeting.

Lamont-Doherty Earth Observatory
Palisades, NY
September 3-5, 2008

Legacy of Ocean Exploration (LOE) project goals –

1.Build central catalog for NSF MG&G-funded cruises

2.Build global multibeam bathymetry synthesis

five-year NSF collaborative project (2005-2010) with LDEO, NGDC, CCOM

Meeting Goals:

1.Review data management current practices on NSF-supported research vessels.

2.Develop options for standard production of underway data + documentation, and delivery to a central repository.

3.Plan next steps and candidates for pilot projects to advance community standards.

Vision

Vessels

Fleet Data Portal

Data Assembly Centers

(SAMOS, MGDS, JASADCP, etc.)

National Data Archives

Benefits:

1. facilitate data preservation and dissemination *(and free operators from archival duty)*
2. standardize + simplify documentation
3. assess + control data quality
4. enable science community to integrate data sets and build visualization apps
5. facilitate cruise planning, permits, clearances

Cruise metadata – application requirements –

- standalone (no network required)
- XML schema-based with versioning and encryption
- forms-based GUI initialized from a template
- provide bonus functionality (e.g. CG manifest)
- email/ftp records to central portal

Vessel profile – instrument metadata –

1. type/class
2. make
3. model
4. serial#
5. firmware
6. install date
7. last calibration date
8. 3D position
9. photo
10. “details”

Event log – application requirements –

- enforce unique event id for given cruise
- auto time stamps
- control vocabulary of event types
- multi-headed (useable throughout ship)
- edits/annotations

Note several ships actively use ELOG.

Underway data – near-realtime datagrams –

- abstracted subset of common types (e.g. nav, met, tsg)
- validate ship's location + activity, plot track map
- central repository should define a fleet standard

dropbox@researchvesseldata.us

Underway data – full distribution post-cruise –

- ftp/rsync to central dropbox or portable drive rotation
- manifest with filenames + checksums
- double-confirm data releases with PIs
- central repository will produce standard products (e.g. “final” navigation, control points, track maps)
- archive must be secure/authenticated

3. Rolling Deck to Repository (R2R) pilot project.

Rolling Deck to Repository (R2R) –

Collaborative 1-year pilot project funded by NSF OCE

Goal: Prototype an end-to-end system to deliver data + documentation from research vessels to a central shoreside repository

Team:

LDEO – R. Arko, V. Ferrini, S. O'Hara

SIO – S. Miller, D. Clark, C. Neiswender

WHOI – C. Chandler, A. Dorsk, S. McCue

Shared development environment –
www.researchvesseldata.us

Implementation –

1. Backend infrastructure
   - Web portal/services
   - SQL database + XML schemas
   - Data store *(SDSC)*
2. Development projects

Development projects –

	Project Description	R2R Lead	Operator Partner
1.	develop vessel profile template, populate for SIO ships	SIO-GDC	F.Delahoyde
2.	establish full data distros from R/V *Kilo Moana* + *Hugh Sharp*	LDEO	T.McGovern, T.Deering
3.	establish near-realtime nav/datagrams from ships	LDEO	“everybody”
4.	develop event logger, test on R/V *Oceanus*	WHOI	A.Dorsk
5.	publish code to generate final nav, test on CGC *Healy* + R/V *Langseth*	LDEO	S.Roberts, A.Johnson
6.	develop best practices for semi-auto q/c of nav+mb, test on SIO ships	SI-GDC	F.Delahoyde
7.	develop data report template, test on R/V *Langseth*	LDEO	A.Johnson

Upcoming events –

- 12/2008 – poster session (IN43) at AGU Fall Meeting
- 02/2009 – working meeting at SIO

Thank you.

info@researchvesseldata.us