Rolling Deck to Repository III: Shipboard Event Logging

AGU Poster IN43A-1170
AGU Fall Meeting
December 17, 2008

http://rvdata.us/

Cynthia L. Chandler[1], Robert A. Arko[2], Stephen P. Miller[3], David C. Fisichella[1], Alexander Dorsk[1], Scott McCue[1]

Rolling Deck to Repository (R2R) Project Overview

NSF-supported research vessels collectively produce an enormous volume and diversity of scientific data. With today's rapidly rising ship costs, and the current trend toward greater re-use of shipboard data, it is imperative that the community takes positive, cost-effective, systematic steps to ensure greater data access.

The NSF Division of Ocean Sciences Data and Sample Policy (pub. NSF 04-004) states, "*Principal Investigators are required to submit all environmental data collected to the designated National Data Centers as soon as possible, but no later than two (2) years after the data are collected. Inventories (metadata) of all marine environmental data collected should be submitted to the designated National Data Centers within sixty (60) days after the observational period/cruise.*" However, procedures for such submissions are poorly established, require lengthy follow-up with investigators, and yield documentation of variable quality. As the volume and diversity of data collected by the fleet increases, this problem will only grow worse.

This new approach provides a "direct pipeline" from operating institutions to a central shoreside facility. Working directly with ship operators, and complementing the data management systems of existing NSF facilities, we will ensure more complete and consistent data collection, quality control, and reporting.

This system will transition the U.S. academic research fleet from a collection of independent expeditionary platforms into an integrated ocean observing system – a network of ships and submersibles around the world that routinely report a standard suite of underway data and documentation to a central repository. The streamlined R2R system will facilitate data discovery and integration, quality assessment, cruise planning, compliance with funding agency data policies, and long-term data preservation.

R2R Poster Series

Rolling Deck to Repository I:
Designing a Database Infrastructure
AGU Poster # IN43A-1168

Rolling Deck to Repository II:
Getting Control of Provenance and Quality
AUG Poster # IN43A-1169

Rolling Deck to Repository III:
Shipboard Event Logging
AGU Poster # IN43A-1170

R2R Project Leads

Woods Hole Oceanographic Institution [1]
Cynthia L. Chandler
cchandler@whoi.edu
Lamont-Doherty Earth Observatory [2]
Robert A. Arko
arko@ldeo.columbia.edu
Scripps Institution of Oceanography [3]
Stephen P. Miller
spmiller@ucsd.edu

The Rolling Deck to Repository Project acknowledges support from the National Science Foundation (NSF), Oceanographic Instrumentation and Technical Services (OITS) Program.

Abstract

Data gathered during NSF-supported scientific research cruises represent an important component of the overall oceanographic data collection. The Rolling Deck to Repository (R2R) pilot project aims to improve access to basic shipboard data and ultimately reduce the work required to provide that access. Improved access will be achieved through identification of best practices for shipboard data management, identification of standard metadata and data products from research cruises, development of metadata schemas to describe a research cruise, and development of a prototype data discovery system that could be used by the entire NSF-supported academic research fleet. Shoreside data managers will work collaboratively with ship technicians and vessel operators to develop approaches that scale from smaller coastal vessels to larger open ocean research vessels.

One of the coordinated subprojects within the R2R project will focus on development of a shipboard event logging system that would incorporate best practice guidelines, a metadata schema and new and existing applications to generate a scientific sampling event log in the form of a digital text file. A cruise event logging system enables researchers to record digitally all scientific sampling events and assign a unique event identifier to each entry. Decades of work conducted within large coordinated ocean research programs (JGOFS, GLOBEC, WOCE and RIDGE) have shown that creation of a shipboard sampling event log can facilitate greatly the subsequent integration of data sets from individual investigators.

In addition to providing a quick way to determine what types of data might have been collected during a cruise, the sampling event log can be used to visualize the relationship, both temporally and geospatially, between the diverse types of sampling events conducted during a research cruise. Research questions in marine ecology or modeling projects are inherently multi-disciplinary and require access to a variety of data types. Improvements in cruise metadata reporting, including a sampling event log that could be contributed routinely to a centralized data repository, should improve access to research cruise data and facilitate accurate reuse of those data by colleagues.

AGU abstract reference number: 12455

Why keep an event log ?

Investigators involved in large coordinated research projects (e.g. JGOFS, GLOBEC, WOCE, RIDGE and Palmer LTER) realized that integration of data from discrete sampling events from all sampling devices deployed during a cruise, was difficult, if not impossible without an accurate record of those sampling events (Baker & Chandler, 2008). For example, if CTD station 4 cast 2 is assigned unique event number 20080904.1342, then the Niskin bottle nutrient data and pigment data analyzed by different investigators can be integrated from that cast using the appropriate event and sample bottle numbers.

What is an event log?

A shipboard scientific event log is a chronological record of all instrument sampling events conducted during a cruise, wherein each sampling event is assigned a unique identifier. The objective is to identify sampling events such that they can be distinguished and described in the relevant X, Y and T dimensions. Sampling event records from an oceanographic research cruise can include (note that [AUTO] means automatic entry from the ship's data system and [CV] means entries come from a controlled vocabulary):

- date and time in Coordinated Universal Time (UTC) and time zone [AUTO]
- longitude and latitude (decimal degrees) [AUTO]
- instrument or sampling device type [CV]
- activity (e.g. deployment, start, max depth, recovery, end, abort) [CV]
- name of person responsible for sampling event [CV]
- a comment field for additional information about the sampling event

If there are discrete sampling events in the Z dimension (height or depth), these are most often recorded in separate device-specific log sheets (e.g. a CTD/Rosette Niskin cast log).

Best Practices

The objective of maintaining a sampling event log is to create an accurate record of all sampling events from a research cruise. This objective can be met in a variety of ways, but experience supporting ocean biogeochemistry cruises and discussions with other data managers indicates several emerging best practice guidelines (BCO-DMO, 2008).

- The event log should be initiated upon ship departure, and concluded upon arrival.
- To minimize errors, event records should be created by on-demand sampling of information from the shipboard data stream.
- Controlled vocabularies should be used instead of free text fields.
- The most accurate position data should be recorded for each event.
- Record date and time in UTC and record time zone to allow calculation of local ship time.
- Include a free text field to capture comments and anecdotal information.
- The contents of the event log should be reviewed regularly during a cruise to ensure completeness and accuracy, and modifications made as soon as possible.
- The final event log should be a comma or tab-separated, plain text, digital file.

mmisw.org Semantic Framework

See the Biological and Chemical Oceanography Data Management Office (BCO-DMO) data system (http://bco-dmo.org) for examples of event logs from several recent ocean research projects.

References

Baker, K.S. and Chandler, C.L., 2008. "Enabling long-term oceanographic research: Changing data practices, information management strategies and informatics. Deep Sea Research Part II, 55 (18-19): 2132-2142.

BCO-DMO, 2008. BCO-DMO Data Management Guidelines Manual, version 1.2 draft. http://bcodmo.org/files/bcodmo/BCO-DMO_best_prac_v1d2.pdf

Buesseler, K.O. and R.S. Lampitt, 2008. Introduction to Understanding the Ocean's Biological Pump: results from VERTIGO. Deep-Sea Research II, 55(14-15): 1519-1521.

Marine Metadata Interoperability (MMI) Semantic Framework, http://marinemetadata.org/semanticframework

NSF, 2004. NSF Division of Ocean Sciences Data and Sample Policy, http://www.nsf.gov/pubs/2004/nsf04004/nsf04004.pdf

UNOLS Data Management Best Practices Committee, http://data.unols.org/

An example from an ocean biogeochemistry research cruise

R/V Kilo Moana
(University of Hawaii Marine Center)

The scientific party on the 2004 VERTIGO cruise aboard R/V Kilo Moana kept an event log to record all sampling events during the research cruise (Buesseler and Lampitt, 2008). The tables below show two different views of the sampling events recorded in the first 6 days of the cruise, the color-colored matrix lists the different types of events with a selection of some of the corresponding event log entries listed in the bottom table.

Year Day	175_176	176_177	177_178	178_179	179_180	180_181
sample day	day 0	day 1	day 2	day 3	day 4	day 5
	23-Jun	24-Jun	25-Jun	26-Jun	27-Jun	28-Jun
hours	4	5	6	7	8	9
0	Trull trap in- 300	Survey CTD #24 & drifters	MOCNESS	MULVFS	NBST 300 out	MOCNESS
2	Launch Clap 150	Survey CTD #25 & drifters	SS	MULVFS	NBST 300 out	NBST 500 out
4	Trull CTD #19 & 3h CTDO2	1000m CTD #26	Trull Rosette #27 (shallow)	MULVFS	CTD #32	NBST 500 out
6	Launch Clap 300	Launch-optical trap	CTD bio-casts #28 (deep)	MULVFS	CTD bio-casts, CTDO2	NBST 500 out
8	Launch Clap 500	Launch-optical trap	MULVFS	NBST 150 out	Clap 300 out	Clap 500 out
10	Launch NBST 150	MOCNESS	MULVFS	NBST 150 out	CTD # 34-39	MOCNESS
12	Launch NBST 300	MOCNESS	MULVFS	NBST 150 out	CTD # 34-39	MOCNESS
14	Launch NBST 500	Launch respirometer	MULVFS	Clap 150 out	CTD # 34-39	CTD # 40-44
16	Survey CTD #21 & drifters	Go-Flo casts	Deep CTD 29- 3000m Bat'th	Clap 150 out	CTD # 34-39	CTD # 40-44
18	Survey & drifters	Go-Flo casts	Deep CTD 29- 3000m Bat'th	CTD #30 & 31	CTD # 34-39	CTD # 40-44
20	Survey CTD #22 & drifters	Launch C explorer	MULVFS	optical trap out	CTD # 34-39	CTD # 40-44
22	Survey CTD #23 & drifters	MOCNESS	MULVFS	bio cast-tow	MOCNESS	CTD # 40-44

Clap Trap

A CTD/Rosette package is recovered.

A sampling device such as a CTD/Rosette package (left) is capable of generating a variety of data sets of interest to researchers. In addition to the unique event identifier associated with each CTD cast, each sample drawn from the Niskin bottles fired at the target depths must be uniquely identified. Unique identification of events and sub-sampling events enables collaborating investigators to more easily integrate disparate data sets.

The 'sample wheel' to the right lists some of the measurements commonly reported from Niskin bottle samples drawn during biogeochemistry research cruises.

event_nmbr	cast	date_UTC	time	lon	lat	person	ev_code	activity	PI_notes_comments
04175_1129	nd	20040623	1129	-157.98	22.78	Trull	Trull_Arr_D1	DEPLOY	Trull Array Dep1
04175_1321	nd	20040623	1321	-157.98	22.78	Trull	CLAP_150_D1D	DEPLOY	CLAP 150m Dep1
04175_1357	19	20040623	1357	-157.97	22.79	Casciotti	CTD_019	DEPLOY	CTD019
04175_1430	19	20040623	1430	-157.97	22.79	Casciotti	CTD_019	MAXDEPTH	CTD019
04175_1555	19	20040623	1555	-157.96	22.79	Casciotti	CTD_019	RECOVER	CTD019
04175_1620	nd	20040623	1620	-157.96	22.79	Buesseler	CLAP_300_D1D	DEPLOY	CLAP 300m Dep1
04175_1915	nd	20040623	1915	-157.94	22.79	Valdes	NBST_13_150_D1D	DEPLOY	NBST #13 150m Dep1
04176_0507	nd	20040624	0507	-158.00	22.76	Silver	HPPT_1	DEPLOY	Hand PP Tow #1
04176_0725	nd	20040624	0725	-157.90	22.85	Fields	SVP2_D1	DEPLOY	SVP2 49659 Dep1
04176_0805	nd	20040624	0805	-157.90	22.85	Fields	SVP2_D1	RECOVER	SVP2 49659 Dep1
04176_0922	nd	20040624	0922	-157.90	22.67	Fields	SVP3_D1	DEPLOY	SVP3 49660 Dep1
04176_1012	nd	20040624	1012	-157.90	22.67	Fields	SVP3_D1	RECOVER	SVP3 49660 Dep1
04177_0245	nd	20040625	0245	-158.02	22.72	Gall	RespArr_D	DEPLOY	Respiration_Array
04178_1715	nd	20040626	1715	-158.07	22.79	Valdes	NBST_13_150_D1R	RECOVER	NBST #13 150m Dep1
04178_1815	nd	20040626	1815	-158.01	22.87	Valdes	CLAP_150_D1R	RECOVER	CLAP 150m Dep1
04178_2200	nd	20040626	2200	-158.24	22.89	Gall	RespArr_R	RECOVER	Respiration_Array
04179_0500	nd	20040627	0500	-158.07	22.89	Bishop	OptTrap_R1	RECOVER	Optical Trap Dep1
04179_1930	nd	20040627	1930	-158.05	22.81	Buesseler	CLAP_300_D1R	RECOVER	CLAP 300m Dep1
04181_1900	nd	20040629	1900	-158.05	22.94	Trull	Trull_Arr_R1	RECOVER	Trull Array Dep1

VERTIGO project sampling gear (images from the VERTIGO project Web site: http://www.whoi.edu/science/MCG/cafethorium/website/projects/vertigo_project.html

Neutrally Buoyant Sediment Trap (NBST)

Multiple Unit Large Volume Filtration System (MULVFS) http://www-ocean.lbl.gov/MULVFS.html

respiration chambers

Plankton Net