NSF Rolling Deck to Repository (R2R) Project Meeting
February 17-19, 2009
Scripps Institution of Oceanography (SIO), La Jolla, CA

Meeting Summary

The R2R pilot project team met at SIO to review progress to date and plan further work. Following is a summary of discussion and consensus from the meeting.

Mission Review

NSF-supported research vessels produce an enormous volume and diversity of scientific data each year. R2R proposes a fundamental transformation of data management in the fleet, providing a "direct pipeline" for the routine underway data and documentation from each cruise to flow from the operating institution to a central shoreside repository. Working directly with the ship scheduling offices and technicians will ensure more complete and consistent data acquisition, quality control, archival, and dissemination.

Cruise Workflow

R2R proposes a set of standard procedures for data stewardship of each cruise. This workflow is expected to evolve continuously, based on field testing and operator feedback.

1. Pre-cruise setup – *rvdata.us* initializes cruise record with basic metadata (vessel, operator, ports, dates, chief scientist, etc) imported from UNOLS Ship Time Request and Scheduling System (STRS); operator logs in to review metadata and populate science party list.
2. Start of cruise – *rvdata.us* automatically transmits cruise record to chief scientist and operator (either technician at sea or manager on shore, depending on vessel); operator initializes R2R standard directory structure on disk including a "dropbox" subdirectory where science party may save e.g. sample logs; confirms metadata and cruise id.
3. Cruise underway – operator provides event log infrastructure and introduction; chief scientist facilitates use of event log by science party.
4. End of cruise – operator populates R2R directory with available standard products (navigation track, event log, cruise-level metadata record, operation report); executes filter to generate checksum file manifest of underway data, and bundles R2R directory with underway data for transmission to *rvdata.us*.
5. Post-cruise check – *rvdata.us* verifies receipt and integrity of data shipment, and archives copy in near-line storage.
6. Data inventory – *rvdata.us* publishes object-level metadata for cruise data shipment – level 1 (file name, size, date, checksum), level 2 (source device and data type classification), level 3 (spatial and temporal extents), and/or level 4 (calibration and type-specific detail).
7. Standard products – *rvdata.us* publishes authoritative objects for every cruise –
    a. Navigation track (cleaned control points)
    b. Scientific sampling event log
    c. Cruise-level metadata per UNOLS Data Committee draft spec. including vessel/instrument profile
    d. Operation report (UNOLS standard style)

8. Quality control – *rvdata.us* performs standard inspection on selected data types e.g. Multibeam (auto-flag bad timestamps and positions, verify sound velocity profiles, etc.), and documents result.
9. Dissemination – resolve proprietary holds if any; deliver routine underway data types to appropriate Data Assembly Centers (DACs) and National Data Centers (NDCs); update cruise catalog record at *rvdata.us* with links to the data sets at those centers.

Underway Data

R2R proposes the "routine underway data types" listed below as its initial scope of work. The list is organized by device (instrument/system) type, indicating the national repositories to which data will be delivered for long-term archival and dissemination.

	Device Type	**Repository**
1.	Navigation (e.g. date/time, position, motion from GPS, INS, VRU, etc.)	R2R
2.	Gravimeter	NGDC
3.	Magnetometer	NGDC
4.	Echosounder (e.g. single/split beam sonar)	NGDC
5.	Multibeam (e.g. swath sonar)	NGDC
6.	Subbottom (e.g. swept-frequency sonar)	NGDC
7.	ADCP (acoustic doppler current profiler)	NODC
8.	CTD (e.g. conductivity, temperature, pressure, dissolved oxygen, pH)	NODC
9.	SSV (sea surface sound velocimeter)	NODC
10.	Expendable Probe (e.g. XBT, XCTD, XSV)	NODC
11.	TSG (thermosalinograph)	NODC
12.	Meteorology Station (e.g. temperature, pressure, humidity, wind speed/direction)	NODC
13.	PAR (photosynthetically available radiation)	NODC
14.	Fluorometer	NODC
15.	pCO2 (partial pressure of carbon dioxide)	NODC
16.	Winch/Wire (e.g. tension, speed, payout)	R2R

Notes:

- Since there is no designated "National Navigation Data Center", R2R will serve as the authoritative repository for such data.
- Data produced by recognized national facilities (MCS, OBSIP, NDSF, etc) or individual investigators with visiting instruments are not classed as routine underway data.
- Some data types will require reduction and quality control beyond what R2R can facilitate. Data Assembly Centers (DACs) that currently provide this function already exist for some data types, e.g. the Academic Seismic Portal (ASP) for Subbottom data and the Shipboard Automated Meteorological and Oceanographic System (SAMOS) for Meteorology and TSG data.
- Other data types such as wildlife (marine mammal/bird) observations merit further discussion.

Data may be delivered to *rvdata.us* in their original format and resolution (raw device outputs or tagged streams from a central acquisition system) as well as "multiplexed" formats (e.g. MGD77/JGOFS files, operator custom formats, etc). The vessel's metadata profile will document the directory/filename convention and format for each of its file types. The long-term goal of R2R is to define a fleet-standard convention.

Fleet Services

R2R will develop and deploy services to benefit the entire academic research fleet community, from vessel operators and schedulers to agency managers and science end users.

- catalog service – *rvdata.us* will publish authoritative cruise-level metadata records with official cruise ids and references to funding awards.
- "dvd rescue" service – operators may bounce requests for lost data sets to *rvdata.us*, where investigators can obtain replacement copies of their original field media.
- clearance service – operators (and U.S. Dept. of State officers) may determine where cruise tracks enter Exclusive Economic Zones, and discover available data types for the spatial/temporal extent in question, as an aid in preparing clearance deliveries to foreign nations.
- "cruise update" service – operators may subscribe to a lightweight (e.g. RSS) service to track a cruise's progress at *rvdata.us* through time (delivery, documentation, and dissemination).
- training guides – *rvdata.us* will publish community best practices for data acquisition, reduction, and quality control (e.g. "how to process gravity data", "known multibeam patch test sites", "how to maintain an event log", etc).
- mapping service – *rvdata.us* will publish standard track maps/products that can be embedded in other Web portals and science applications.

R2R will actively solicit community feedback and engage operators in prototype development. Suggested venues include a *@rvdata.us* Web portal and mailing list, engagement at relevant meetings (UNOLS RVTEC, AGU, INMARTECH, etc), appointment of a R2R Advisory Board, and publication in relevant journals (*EOS*, *Sea Technology*).