Shipboard Data Stewardship *from Ship to Shore to Archive*

Cynthia L. Chandler[1], Robert A. Arko[3], M. Dickson Allison[2], Tobias Work[2], Scott Mowery[4]
cchandler@whoi.edu, arko@ldeo.columbia.edu, dallison@whoi.edu, twork@whoi.edu, Scott.Mowery@noaa.gov

1~Marine Chemistry and Geochemistry Department, 2~Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA
3~, Lamont-Doherty Earth Observatory (LDEO), Columbia University, 61 Route 9W, Palisades, NY 10964 USA
4~National Oceanic and Atmospheric Administration National Oceanographic Data Center (NOAA NODC), Silver Spring, MD 20910 USA

US National Science Foundation

Rolling Deck to Repository (R2R)
rvdata.us

BCO-DMO
Biological and Chemical Oceanography Data Management Office

Archives, WHOI

Introduction

The increased volume and heterogeneity of scientific research data present new challenges for data managers tasked with the stewardship of those valuable data resources. Stewardship of *in situ* oceanographic data is especially important as those data form an essential part of the time-series observations comprising the full climate data set for planet Earth. Acquisition of oceanographic data by U.S. funded investigators often begins aboard a vessel in the U.S. academic fleet and data are managed subsequently by individual investigators with the understanding that all data will be archived at the appropriate U.S. National Oceanic and Atmospheric Administration (NOAA) National Data Center. In late 2006, the U.S. National Science Foundation (NSF) funded the Biological and Chemical Oceanography Data Management Office (BCO-DMO) at Woods Hole Oceanographic Institution to work closely with investigators to manage data generated from their research projects and to ensure that those data are permanently archived at the U.S. National Oceanographic Data Center (NODC). The Rolling Deck to Repository (R2R) project, funded by NSF in late 2009, has been tasked to work with vessel operators to improve stewardship of shipboard data routinely collected from suites of underway sensor systems and ensure final archive of those data at the appropriate U.S. National Data Center, e.g. the NODC or the National Geophysical Data Center (NGDC).

Chris Linder, WHOI

The Old Way

Data managers hunted for cruise metadata and attempted to assemble the complete metadata record for each cruise.

NSI/DECnet email 301-286-9790 NSF::MORRISON

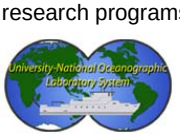
CRUISE METADATA

vessel name
cruise ID
Chief Scientist name
Sponsoring Institution
Project Name
Dates

ONLINE DATA SYSTEM

A Better Way

Data managers collaborate and agree on the use of established international vocabularies, metadata content standards, and Web services in developing data management procedures that scale to support the increased volume and complexity of data generated by modern oceanographic research programs.



UNOLS ship scheduling system

RESEARCH VESSELS

CRUISE METADATA

Vessel (ICES code + full name)
Cruise ID (R2R authoritative code)
Chief Scientist Name (R2R name)
Sponsoring Institution (IANA institution code)
Project Name
Start and End Dates (ISO)
Primary Funding (IANA agency code + award id)
Start and End Port (UNOLS code + full name)
Cruise track control points (SFS geometry)

R2R provides a basic ("brief") cruise-level metadata record that contains the elements listed above, via a lightweight ReST / WADL style search service.

BCO-DMO consumes the service and populates the appropriate fields in the metadata database.

Platform table
- platform_name (vessel)
- deployment (cruise ID)
- start_date
- end_date

Project table
- project_name

Project_funding table
- funding_id (from Funding table)

People table
- first_name
- last_name

Person_role table
- title_lookup_id (= Chief_Scientist)

assessment + feedback

data + documentation

R2R

data + documentation

unified catalog + standard cruise products

DATA ASSEMBLY CENTERS

integration + synthesis

NATIONAL DATA CENTERS

archiving + dissemination

DATA ASSEMBLY CENTER

BCO-DMO
Biological and Chemical Oceanography Data Management Office

push
HTML + CSV / TXT (via FTP) + Email notification

push
XML + CSV / TXT (via FTP) or JSON (via ReST)

NOW

FUTURE

Accession Number (via Email)

Accession Number (via POST) + Persistent Identifier (DOI)

NODC
National Oceanographic Data Center

NATIONAL DATA CENTER

OPEN ACCESS TO DATA