Rolling Deck to Repository: Linking Data Across the Research Fleet

Curators Meeting - April 25-27, 2017 (Corvallis)
Mission

Stewardship of environmental sensor data routinely acquired by U.S. academic research vessels

Services:
• Publish master cruise catalog
• Document, preserve, disseminate original field datasets
• Assess quality of selected data types
• Create post-field data products
• Support at-sea event logging

Supported by: NSF, ONR, NOAA, SOI, UNOLS
Sensor Types

- Geophysics:
  - Echosounder
  - Gravimeter
  - Magnetometer
  - Multibeam
- Meteorology:
  - Anemometer
  - Barometer
  - Hygrometer
  - Radiometer
  - Rain Gauge
  - Temperature
- Oceanography:
  - ADCP
  - CTD
  - Fluorometer
- Navigation:
  - Camera
  - GNSS/INS
  - Gyrocompass
  - Speedlog
Partnerships

Growing number of DataONE nodes that hold marine data:

- NCEI
- BCO-DMO
- LTER Sites
- AOOS
- IARC
- R2R

NSF - Supported Federation of Data Repositories
Partnerships

SeaView project: Create thematic data collections, organized+formatted.

Collaborators
CCHDO
BCO-DMO
* R2R
OOI
OBIS

K. Stocks, Lead

https://eos.org/articles/unifying-ocean-data-into-one-searchable-set
Cruise Catalog

- Vessel ID
- Cruise ID+Title
- Trackmap
- Program,Campaign
- Dates+Ports
- Science Party
- File Manifest
- Underway Data
- QA Reports
- Products
- Related Cruises
- Publications
- Funding
- Related Data
Cruise Catalog

Cruise

ScienceDirect

Journal Article

Samples

Researcher Profile

NOAA National Centers for Environmental Information

Index to Marine and Lacustrine Geological Samples (IMUGS)

44 Samples/Holes from MGL1208 are in the Index

<table>
<thead>
<tr>
<th>Repository</th>
<th>Core/Platform</th>
<th>MSI Data Link</th>
<th>Core/Gravity</th>
<th>Device</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Water Depth (m)</th>
<th>Collection Date</th>
<th>ISDN Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDNO</td>
<td>Marcus G. Langseth</td>
<td>MGL1208</td>
<td>MGL1208-02</td>
<td>core, gravity</td>
<td>8.2094</td>
<td>-180.19213</td>
<td>1349</td>
<td>20120430</td>
<td>DSR00054E</td>
</tr>
<tr>
<td>LDNO</td>
<td>Marcus G. Langseth</td>
<td>MGL1208</td>
<td>MGL1208-03</td>
<td>core, gravity</td>
<td>6.57087</td>
<td>-161.23856</td>
<td>2088</td>
<td>20120430</td>
<td>DSR00054F</td>
</tr>
<tr>
<td>LDNO</td>
<td>Marcus G. Langseth</td>
<td>MGL1208</td>
<td>MGL1208-04</td>
<td>core, gravity</td>
<td>6.3405</td>
<td>-161.27333</td>
<td>1803</td>
<td>20120430</td>
<td>DSR00054G</td>
</tr>
<tr>
<td>LDNO</td>
<td>Marcus G. Langseth</td>
<td>MGL1208</td>
<td>MGL1208-05</td>
<td>core, gravity</td>
<td>6.3405</td>
<td>-161.27333</td>
<td>1803</td>
<td>20120430</td>
<td>DSR00054H</td>
</tr>
<tr>
<td>LDNO</td>
<td>Marcus G. Langseth</td>
<td>MGL1208</td>
<td>MGL1208-06</td>
<td>core, gravity</td>
<td>6.408</td>
<td>-161.0075</td>
<td>2371</td>
<td>20120430</td>
<td>DSR00054I</td>
</tr>
<tr>
<td>LDNO</td>
<td>Marcus G. Langseth</td>
<td>MGL1208</td>
<td>MGL1208-07</td>
<td>core, gravity</td>
<td>6.35913</td>
<td>-160.7886</td>
<td>3163</td>
<td>20120430</td>
<td>DSR00054J</td>
</tr>
<tr>
<td>LDNO</td>
<td>Marcus G. Langseth</td>
<td>MGL1208</td>
<td>MGL1208-08</td>
<td>core, piston</td>
<td>6.368</td>
<td>-160.7887</td>
<td>3163</td>
<td>20120430</td>
<td>DSR00054K</td>
</tr>
<tr>
<td>LDNO</td>
<td>Marcus G. Langseth</td>
<td>MGL1208</td>
<td>MGL1208-09</td>
<td>core, trigger</td>
<td>6.358</td>
<td>-160.7872</td>
<td>3163</td>
<td>20120430</td>
<td>DSR00054L</td>
</tr>
</tbody>
</table>

(cont.)
Cruise Search

- Vessel
- Cruise ID
- Dates
- Program
- Campaign
- Partner
- WESN
- Institution
- Person

http://www.rvdata.us/search
Data Syntheses

Key metric of success is volume of data contributed to global synthesis products.

Current:

- Global Multi-Resolution Topography (GMRT)
- Int’l Comprehensive Ocean-Atmosphere Data Set (ICOADS)

On Deck:

- NCEI Multibeam Bathymetry Mosaic/Subsets
- World Ocean Database - CTD/XBT
- Global Ocean Currents Database (GOCD)
- Global Thermosalinograph (TSG) Synthesis
Data Syntheses

Global Multi-Resolution Topography (GMRT) Synthesis

Above: Shiptracks for Multibeam data contributed by R2R (202 cruises as of June 2016).

GMRT produced by the Interdisciplinary Earth Data Alliance (IEDA)
The Future: Persistent Identifiers

- PIDS for Cruises + Datasets
- PIDS for Organizations
- PIDS for Researchers
- PIDS for Samples + Features
- PIDS for Articles + Awards

DataCite, ORCID, IGSN, Crossref
Persistent Identifiers (cont.)

Motivation: The 3 R’s

- Reproducibility - what journals want
- Reuse - what funders want
- Recognition - what scientists want
Persistent Identifiers (cont.)

Example “Cruise” DOI for MGL1208

Repository # for R2R
ORCID of CS
OFR # for NSF
DOI of a paper

Next step: Add IGSNs
Persistent Identifiers

• Field Deployments
  • Ocean Observing Initiative (OOI) will reuse R2R’s Cruise DOIs
  • Int’l Ocean Discovery Program (IODP) will publish DOIs for expeditions
  • Continental Drilling Office (CSDCO) will publish DOIs for projects

• Persons and Organizations
  • R2R has mapped 740+ ORCIDs and 540+ GRID IDs
Persistent Identifiers (cont.)

ORCID

Cruise @R2R

Award @NSF

Paper

Dataset @BCO-DMO

Dataset @ECL

Samples @SESAR
Next Steps

1. Work with Repositories (both Samples and Data) to enhance their standard operating procedures:
   - Link Samples to Deployment IDs
     - R2R Cruises
     - IODP Expeditions
     - CSDCO Projects
     (etc.)
   - Link Datasets to Sample IDs
   - Links Researchers to ORCID IDs

2. Work with Publishers to link Datasets +Samples to Papers