

Rolling Deck to Repository (R2R): Lessons Learned

R.Arko¹, S.Carbotte¹, C.Chandler², S.Smith³, K.Stocks⁴, L.Stolp²

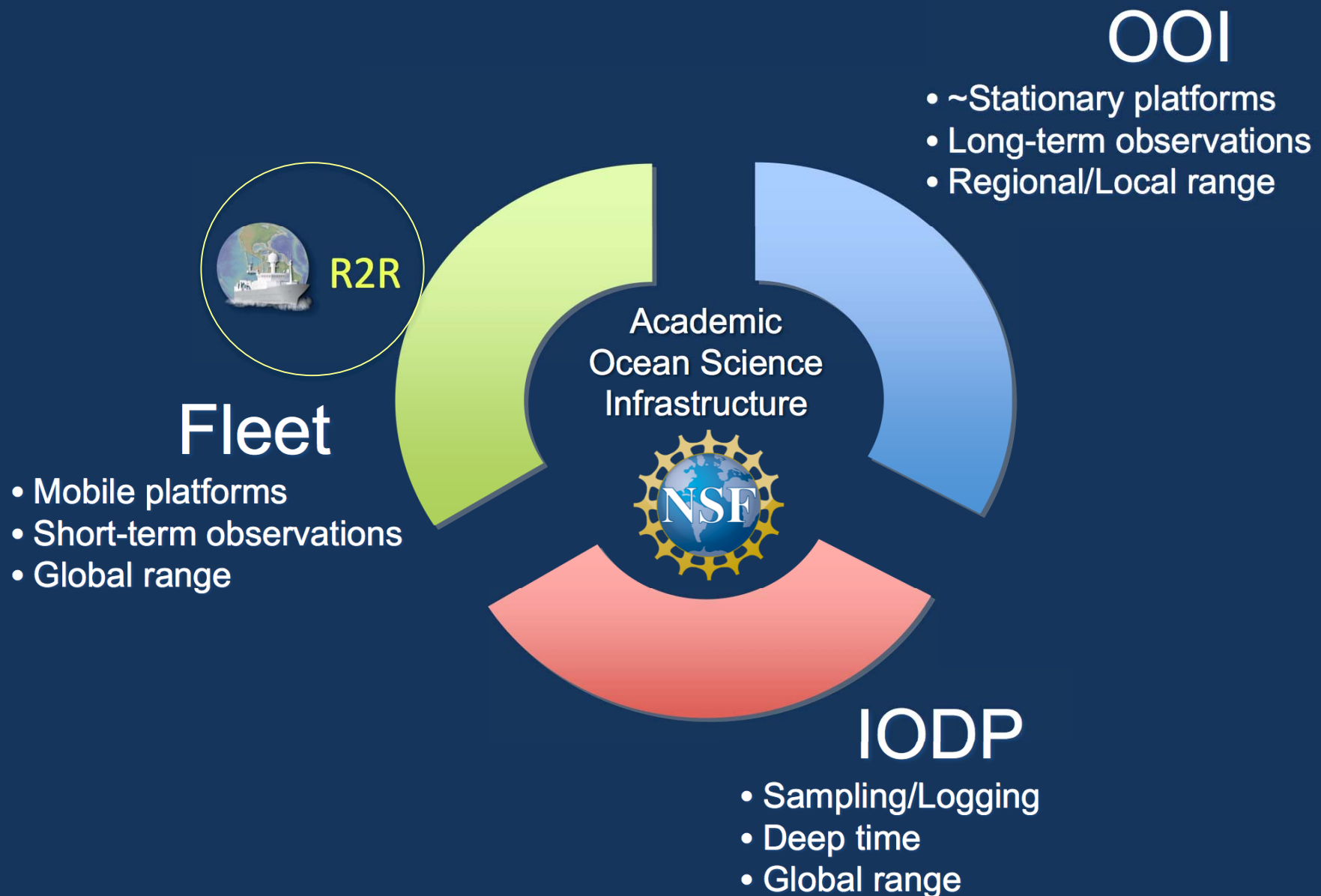
¹LDEO ²WHOI ³FSU ⁴SIO



Part#1: Overview



Context



Overview

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

Support:

NSF , ONR, NOAA, SOI, UNOLS



Alpha Helix (retired)



Atlantic Explorer



Atlantis



Blue Heron



Cape Hatteras (retired)



Clifford A. Barnes



Corwith Cramer



Endeavor



Eitanin (retired)



F.G. Walton Smith



Falkor



Healy



Hugh R. Sharp



Ka`imikai-O-Kanaloa



Kilo Moana



Knorr (retired)



Laurence M. Gould



Marcus G. Langseth



Maurice Ewing (retired)



Melville (retired)



Nathaniel B. Palmer



Neil Armstrong



New Horizon (retired)



Oceanus



Okeanos Explorer



Pelican



Point Sur



Polar Sea



Robert C. Seamans



Robert D. Conrad (retired)



Robert Gordon Sproul



Roger Revelle



Ronald H. Brown



Sally Ride



Savannah



Sikuliaq



Thomas G. Thompson



Vema (retired)



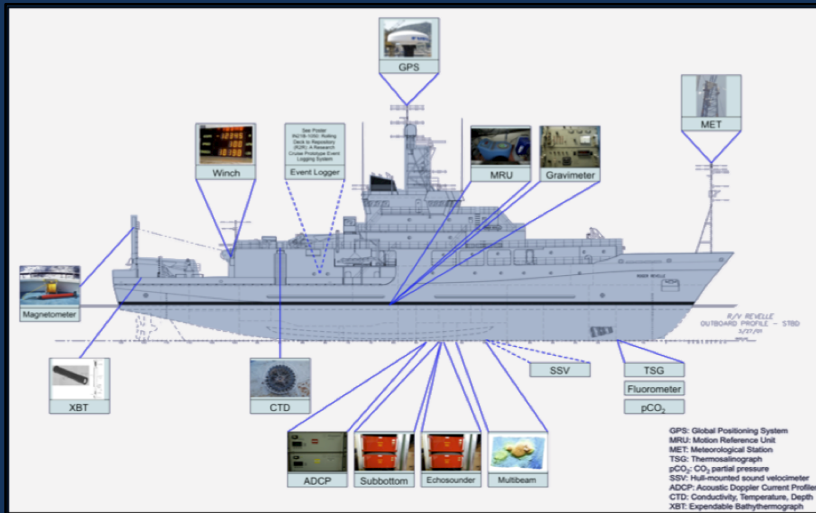
Thomas Washington (retired)



Wecoma (retired)



Sensor Types



Geophysics:

Echosounder
Gravimeter
Magnetometer
Multibeam

Meteorology:

Anemometer
Barometer
Hygrometer
Radiometer
Rain Gauge
Temperature

Oceanography:

ADCP
CTD
Fluorometer

:

Nutrients
Oxygen
pCO2
pH
Temperature
Transmissometer
TSG
Wave Radar
XBT/XCTD

Navigation:

Camera
GNSS/INS
Gyrocompass
Speedlog

:



www.rvdata.us/catalog/EN537


Supported by: NSF | NOAA | ONR | SOI

Rolling Deck to Repository (R2R)

Home About R2R Cruise Catalog QA Dashboard News Contact Us Internal

Catalog Status
 (In Service) Vessels: 22
 Cruises: 6045
 Archived Files: 22216407
 April 4, 2016

Home
 Cruise Catalog: EN537



Operator: University of Rhode Island
 Vessel: Endeavor
 Cruise DOI: 10.7284/904058

Cruise ID	Start Date	Start Port	End Date	End Port
EN537	2014-04-01	Narragansett, Rhode Island	2014-04-16	Narragansett, Rhode Island

Details

Project: Eastern North American Margin (ENAM) Community Seismic Experiment (Info)

SCIENCE PARTY

FILE MANIFEST

UNDERWAY DATA SETS (ORIG FIELD DATA)

Device Type	Make-Model [Location]	Files	Archive Status
adcp	Hawaii UHDAS	List	R2R Download
ctd	Sea-Bird SBE-911plus	List	R2R Download
fluorometer	Turner 10-AU	List	R2R Download
gnss	Furuno GP-90D	List	R2R Download
gnss	Northstar 952	List	R2R Download
gnss	Trimble NT-300D	List	R2R Download
gyrocompass	Sperry MK-37 [2]	List	R2R Download
gyrocompass	Sperry MK-37 [1]	List	R2R Download
speedlog	EDO MRQ-4015D	List	R2R Download
tsg	Sea-Bird SBE-21	List	R2R Download

R2R PRODUCTS (POST-FIELD PROCESSED)

RELATED CRUISES

EN546

MGL1408

FUNDING AWARDS

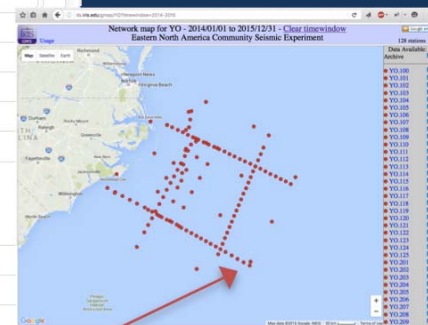
RELATED DATA

IRIS Data Management Center (Info)

Data (OBS)

Cruise Catalog

- Vessel ID
 - Cruise ID +Title
 - Trackmap
 - Program,Campaign
 - Dates/Ports
 - Science Party
 - File Manifest
 - Datasets
 - Standard Products
 - Funding
 - Related:
- Cruises
 Datasets
 Articles



Cruise Search

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

Cruise Search

Vessel: (ANY)

Cruise: (quick select of recent cruises in the last ~three years) (ANY)

Dates: 1950 01 01 From 2020 12 31 To

Program: (broad community initiative, typically decadal-scale) **Cascadia**

Campaign: (series of closely-related cruises for a single project) (ANY)

Location: (pan and zoom to adjust bounding box)

N: 88
W: -180 E: 180
S: -76

Person: (science party and/or vessel technicians) (ANY)

Institution: (ANY)

Search Reset

Search Result

Query = Cascadia Initiative (Cascadia)
Number of Cruises = 25

Cruise ID	DOI	Vessel	Depart Date	Port	Arrive Date	Port
TN331	10.7284/906264	Thompson	2015-10-01	Newport	2015-10-15	Newport
OC1508B	10.7284/906267	Oceanus	2015-08-26	Port Hueneme	2015-09-02	Newport
OC1409A	10.7284/900721	Oceanus	2014-09-07	Newport	2014-09-26	Newport
OC1408A	10.7284/900719	Oceanus	2014-08-09	Newport	2014-08-13	Newport
OC1407B	10.7284/900718	Oceanus	2014-07-24	Newport	2014-08-06	Newport
OC1407A	10.7284/900717	Oceanus	2014-07-10	Newport	2014-07-19	Newport
TN312	10.7284/900991	Thompson	2014-06-22	Newport	2014-07-06	Seattle
OC1405B	10.7284/900714	Oceanus	2014-05-28	Newport	2014-06-02	Newport
OC1405A	10.7284/900713	Oceanus	2014-05-13	Newport	2014-05-23	Newport
OC1308B	10.7284/900694	Oceanus	2013-08-29	Newport	2013-09-06	Newport
OC1308A	10.7284/900693	Oceanus	2013-08-18	Newport	2013-08-22	Newport
OC1307C	10.7284/901918	Oceanus	2013-08-01	Newport	2013-08-10	Newport
AT26-02	10.7284/900486	Atlantis	2013-06-25	Astoria	2013-07-09	Astoria
OC1306A	10.7284/901805	Oceanus	2013-06-17	Newport	2013-06-22	Newport
OC1305C	10.7284/901804	Oceanus	2013-06-03	Newport	2013-06-14	Newport
OC1209A	10.7284/900969	Oceanus	2012-09-10	Newport	2012-09-20	Newport
OC1208B	10.7284/900683	Oceanus	2012-08-31	Newport	2012-09-07	Newport
OC1208A	10.7284/902582	Oceanus	2012-08-23	Newport	2012-08-30	Newport

<http://www.rvdata.us/search>



Shiptrack Navigation

R2R has developed open-source software package to:

- Convert original GNSS/INS data to a common format
- Perform quality assessment
- Produce suite of down-sampled quality-controlled files

<http://github.com/rvdata/NavManager>



Quality Assessment

- GNSS/INS (navigation)
- METOC (marine meteorology)
- Multibeam
- Gravity
- Magnetics
- Subbottom *New*
- CTD *New* (water column profiles)
- ADCP *(2017)* (ocean currents)
- Flowthru *(2017)*



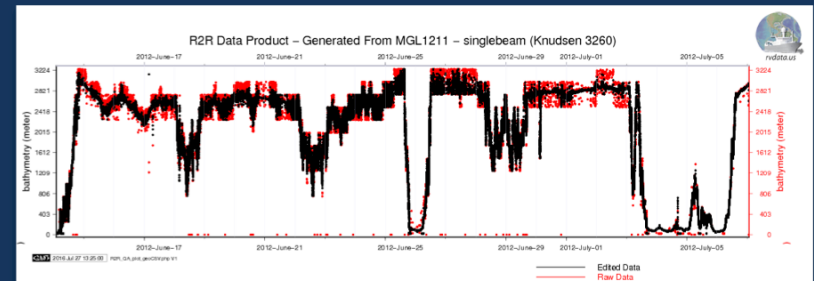
(cont.)

Quality Assessment

Example: Echosounder data

Typical Tests:

- % NonZeroLengthFiles
- % ValidFormatFiles
- % RecordsWithoutErrorCode
- % RecordsWithValidDepth
- % RecordsWithValidTime
- % RecordsWithValidPosition
- % RecordsWithValidPositionFormatFlag
- % RecordsWithoutOutliers
- % FilesWithSoundVelocity
- % FilesWithSubbottom
- HasRequiredMetadata



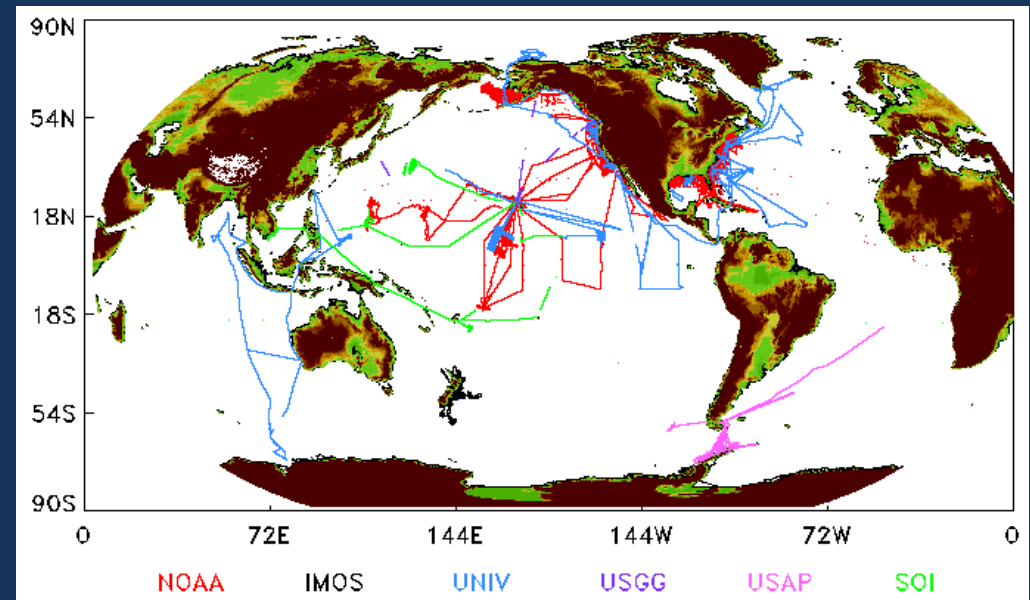
*Edited bathy profile created where possible.
(Example: MGL1211)*





SAMOS Data Assembly Center

- Produce quality-controlled METOC data in near-real-time
- 31 research vessels participating in FY2016
- ~30-40K 1-min. observations per month per vessel



(SAMOS data coverage: Oct 1, 2015 - Sep 30, 2016)



Event Logger

Recent deployments
Jun. 2015-May 2016:



(DreamPlug™)

<i>Armstrong</i>	AR1-01	Trials
	AR1-03	Trials
	AR1-06	Corbett
	AR1-07	Lin
	AR3	Shank
	AR4B,C	Plueddemann
<i>Atlantis</i>	AT31	Plueddemann
	AT31B,C	Plueddemann
	AT32	Behrenfeld
	AT33-02	Equip. Test
	AT33-03	Kurz
	AT34	Behrenfeld
<i>Revelle</i>	RR1508	Harris
	RR1601	Appelgate
	RR1603	McDonald
	RR1604	Barbero
<i>Sikuliaq</i>	SKQ201512S	Thompson
<i>Thompson</i>	TN326	Kelley
	TN342	Fram



Part#2: Lessons



#1. Motivation

The 3 “R”s

- Reproducibility (what journals want)
- Reuse (what funders want)
- Recognition (what scientists want)



#1. Motivation

(cont.)

ORCID

Cruise @R2R

Award @NSF

Paper

Report
Aerobic Microbial Respiration in 86-Million-Year-Old Deep-Sea Red Clay
Hans Rey^{1,*}, Jens Kallmeyer², Rishi Ram Adhikari², Robert Pockalny², Bo Barker Jørgensen¹, Steven D'Hondt¹
*To whom correspondence should be addressed. E-mail: hans.rey@biology.aau.dk
Science 18 May 2012:
Vol. 336, Issue 6083, pp. 922-925
DOI: 10.1126/science.1219424
Article
Figures & Data

ORCID
Jens Kallmeyer
ORCID ID
orcid.org/0000-0002-6440-1140
Other IDs
Loop profile: 30680
ResearcherID: I-3554-2012
Education (4)
Max-Planck-Institute for Marine Microbiology: Bremen, Bremen, Germany
2002-02 to 2003-05-30 (Biogeochemistry)

Rolling Deck to Repository (R2R)
Supported by: NSF | NOAA | ONR | SOI
Home About R2R+ Cruise Catalog QA Dashboard News Contact Us Internal
Cruise Catalog: KN195-03
Operator: Woods Hole Oceanographic Institution
Vessel (retired): Knorr
Cruise DOI: 10.7284/V195-03

National Science Foundation
WHERE DISCOVERIES BEGIN
Award Abstract #0752336
Oceanographic control and global distributions of subsurface microbial life and activity
NSF Org: OCE
Division Of Ocean Sciences
Initial Amendment Date: March 21, 2008
Latest Amendment Date: March 21, 2008
Award Number: 0752336
Award Instrument: Standard Grant
Program Manager: David L. Garrison
OCE Division Of Ocean Sciences
GEO Directorate For Geosciences
Start Date: September 1, 2008
End Date: August 31, 2013 (Estimated)

EarthChem Library Record
Dataset Information
Basic Information
Dataset Title: Dissolved inorganic carbon measurements of marine sediment from Long Core Expedition KN195-03
Author(s): Savage, Justin; D'Hondt, Steven; Pockalny, Robert
Dataset DOI: doi:10.1594/IEDA/100942
Date Released/Published: 10/15/2015
Abstract or Description: The KN195-03 research cruise on board the R/V Knorr (Chief scientist: Steven D'Hondt) took place in January-February 2009 and had as primary objective to test and refine quantitative models for the magnitude and geographic distribution of subsurface biomass and organic-fueled subsurface respiration. Sediment cores were taken at thirteen sites across the equatorial Pacific and the North Pacific Gyre. Knowledge of the chemical composition of the porewater throughout the collected cores is fundamental to answer the driving research questions of this scientific expedition. The shipboard geochemistry laboratory generated interstitial waters by sparging 397 individual samples, and by extracting an additional 133 sediments by Rhizon sampling. Due to the time needed to analyze this very high number of samples, not all were analyzed for each dissolved constituent. We present here the shipboard porewater dissolved inorganic carbon concentration dataset measured by Marianda AIRCA™/Infared system.
Citation: Savage, Justin; D'Hondt, Steven; Pockalny, Robert (2015): Dissolved inorganic carbon measurements of marine sediment from Long Core Expedition KN195-03. EarthChem Library. <http://dx.doi.org/10.1594/IEDA/100942>

SESAR
Sample Search
Samples Found: 795
Table with columns: ID#, Sample Name, Object Type, Material Classification, Latitude, Longitude, Location

BCO-DMO
Biological and Chemical Oceanography Data Management Office
Dataset: Subseafloor Microbial Cell Counts
Deployment: KN195-03
Microbial cell counts in sediment cores collected during KN195-03.
Principal Investigator: Dr Steven L D'Hondt (University of Rhode Island, URI-GSO)
Co-Principal Investigator: Dr Robert Pockalny (University of Rhode Island, URI-GSO)
Dr David C. Smith (University of Rhode Island, URI-GSO)
Dr Arthur J. Spivack (University of Rhode Island, URI-GSO)
Contact: Dr Jens Kallmeyer (Helmholtz Centre Potsdam, GFZ)
BCO-DMO Data Manager: Shannon Rauch (Woods Hole Oceanographic Institution, WHOI BCO-DMO)
Project: Oceanographic control and global distributions of subsurface microbial life and activity (Subseafloor Microbial Life)

Dataset @ECL

Samples @SESAR

Dataset @BCO-DMO



#2. Federation

R2R joined DataONE to:

- Make content more discoverable
- Find collaborators
- Access user-facing apps eg. Investigator Toolkit



Growing number of DataONE nodes that hold marine data:



NCEI



LTER Sites



R2R

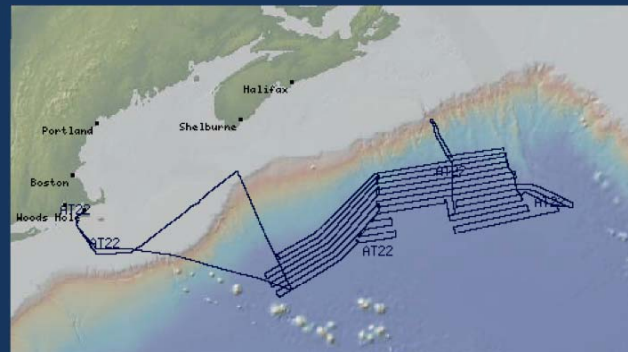
⋮



#3. Information model

In R2R, a “Cruise” is ..

- ... a deployment of a known research platform
- ... starting and stopping at named locations (port calls)
- ... where position is continuously recorded
- ... with scientists and/or technicians onboard
- ... whose purpose is one of
 - Research (funded mission to collect data +samples)
 - Transit
 - Maintenance



*R/V Atlantis cruise
AT22 (Scotian Shelf
Survey, 2012)*



#3. Information model ^(cont.)

In R2R, a “Dataset” is ..

- ... a collection of digital objects
- ... produced by a single instrument system (type/make/model)
- ... in a known file format
- ... recorded by an acquisition system on a defined interface
- ... during a single Cruise



#3. Information model ^(cont.)

In R2R, a Dataset may have relationships with other Datasets:

1. Dataset Y IsNewVersionOf Dataset X
if Y is a corrected version of X that replaces missing or corrupt files.
2. Dataset Y IsIdenticalTo Dataset X
if Y is an exact copy of X that was published by another repository.
3. Dataset Y IsDerivedFrom Dataset X
if Y is a new set of files that was created by processing or editing X.



#4. Citation

R2R publishes a DOI for every Dataset as soon as it is released and published online ie. downloadable.

DOI metadata record describes the Dataset quality.

Quality should not be “inferred” from the absence/presence of a DOI.

UNIVERSITY OF CALIFORNIA **EZID** Identifiers made easy

SEARCH LEARN CONTACT LOGIN

Identifier Details

ABOUT THE IDENTIFIER

Identifier: doi:10.7284/123707
Identifier as URL:

ABOUT THE IDENTIFIED OBJECT

Location (URL): <http://data.rvdata.us/id/fileset/123707>
Creators: Rolling Deck to Repository [DOI=10.17616/R39C8D]
Title: Gravimeter data collected during research cruise AT37-05 using a Bell BGM-3 instrument system onboard the platform RV Atlantis
Publisher: Rolling Deck to Repository (R2R) Program
Publication year: 2016
Language: en
Resource type: Dataset/Environmental
Description [Methods]: Original (unprocessed) data collected during a field expedition
Related identifiers: 10.7284/907107 [DOI] [IsPartOf]
Version: 1



#5. Vocabularies

A vocabulary qualifies for reuse if ...

- ... it is published on the Web

- ... with URIs that can be dereferenced

- ... that return RDF (Resource Description Framework) format

eg. SKOS or OWL

- ... every term has a definition

- ... and it has community governance

R2R reuses existing vocabularies for both

Classes (Types)

eg. “Research Vessels”

and

Instances

eg. “*R/V Hugh R. Sharp*”



(cont.)

#5. Vocabularies

R2R uses terms from the NERC Vocabulary Server:

Classes

L06 Platform Types

L05 Instrument Types

L22 Instrument SubTypes (Make/Model)

C77 Shipboard Activity Types

P02 Data Types

P06 Unit Types

Instances

C17 Platforms

C19 Waterbodies

C32 Countries

C38 Ports

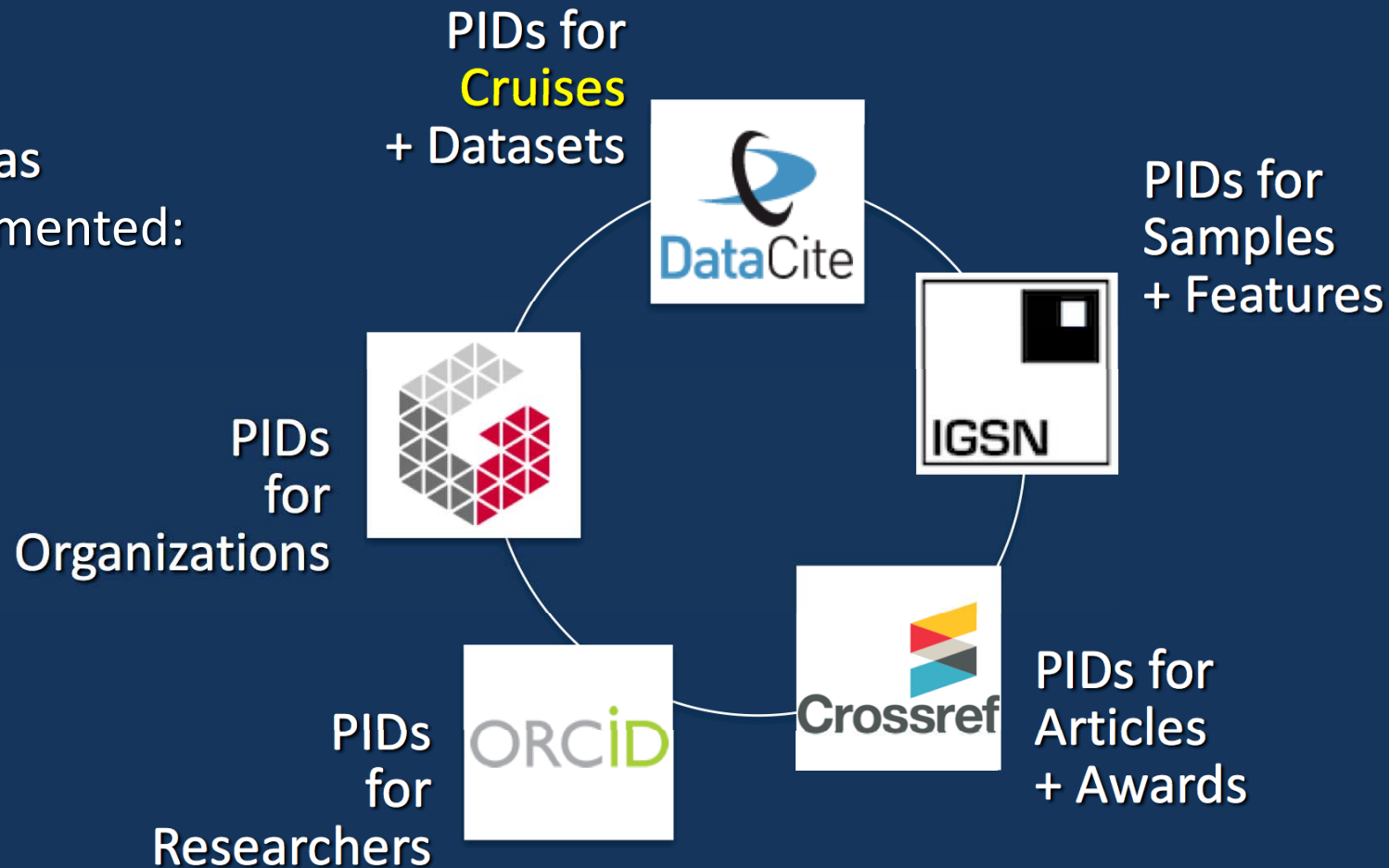


<http://vocab.nerc.ac.uk/>



#6. Persistent Identifiers

R2R has implemented:



After: Hanson, B. (2016), AGU opens its journals to author identifiers, *Eos* 97, doi:10.1029/2016EO043183.



(cont.)

#6. Persistent Identifiers

Challenges with adopting persistent identifiers:

- Community has not reached consensus on all identifier types eg. Instruments, Projects, Organizations.
- Significant backfill mapping required.
- Identifier systems are not perfect, eg. a person can have more than one ORCID.



#7. Partnerships

R2R has “reciprocal linking” arrangements with other repositories eg.

- Biological and Chemical Oceanography Data Management Office (BCO-DMO)
- CLIVAR and Carbon Hydrographic Data Office (CCHDO)
- Interdisciplinary Earth Data Alliance (IEDA)
- Index to Marine and Lacustrine Geological Samples (IMLGS)
- Incorporated Research Institutions for Seismology (IRIS/OBSIP)
- Shipboard Automated Meteorological and Oceanographic System (SAMOS)
- System for Earth Sample Registration (SESAR)

The image shows two screenshots of web pages. The top screenshot is from the BCO-DMO website, displaying the details for deployment KN195-03. A red circle highlights the text: "Cruise information and original data are available from the NSF R2R data catz". The bottom screenshot is from the R2R website, showing a map of the cruise track in the Pacific Ocean. A red circle highlights a link labeled "Bio+Chem Data". A red line connects the two circles, indicating the reciprocal linking between the two repositories.

BCO-DMO Deployment: KN195-03

Deployment: KN195-03
Chief Scientist: Dr Steven L. D'Hondt (University of Rhode Island, URI-GSO)
Synonyms: EQP
Equatorial Pacific
Platform: R/V Knorr
Platform Type: vessel
Start Date: 2009-01-12
End Date: 2009-02-23
Location: Pacific Ocean

Description
The cruise went from Puntarenas, Costa Rica to Honolulu, Hawaii and was the used the long coring system (developed at Woods Hole Oceanographic Institution seafloor).
Cruise information and original data are available from the NSF R2R data catz

R2R Catalog: KN195-03

Operator: Woods Hole Oceanographic Institution
Vessel (retired): Knorr
Cruise DOI: 10.7284/900496

Cruise ID	Start Date	Start Port	End Date	End Port
KN195-03	2009-01-12	Puntarenas, Costa Rica	2009-02-23	Honolulu, Hawaii

Project: Oceanographic Control and Global Distributions of Subseafloor Microbial Life and Activity

- SCIENCE PARTY
- FILE MANIFEST
- UNDERWAY DATA SETS (ORIGINAL FIELD DATA)
- R2R PRODUCTS (POST-FIELD PROCESSED)
- FUNDING AWARDS
- PUBLICATIONS
- RELATED DATA

Biological and Chemical Oceanography Data Management Office (Info) [Bio+Chem Data](#)
Index to Marine and Lacustrine Geological Samples (Info) [Samples](#)
EarthChem Library (Info) [Geochem Dataset#898](#)



(cont.)

#7. Partnerships



R2R collaborates in cross-cutting EarthCube projects eg.

SeaView - integrating data from multiple repositories to create regional/thematic collections

X-DOMES - working with vendors to create standard instrument descriptions

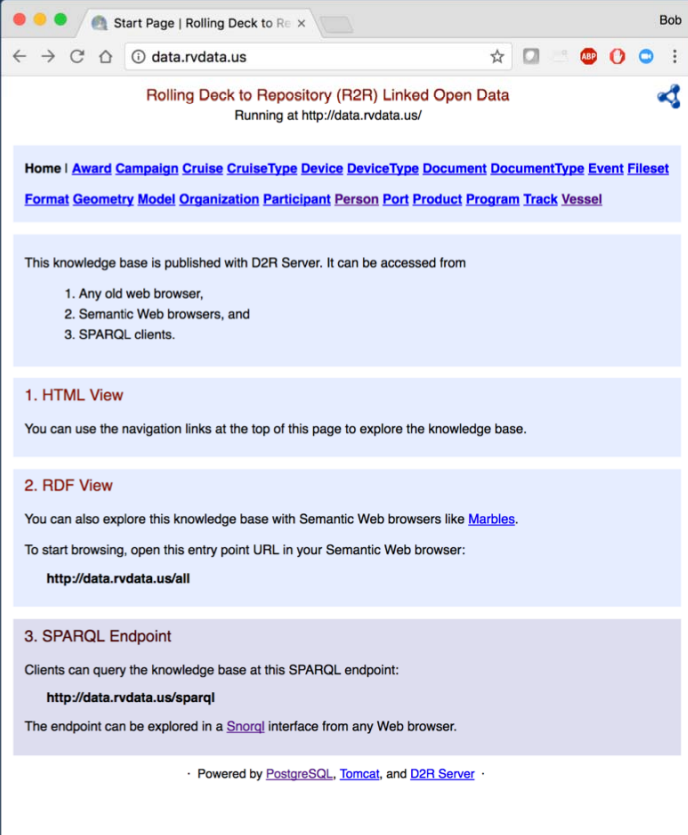
GeoLink - publishing resources on the Web as Linked Data and developing shared semantics



#8. Reliability

Partner repositories rely on R2R's Cruise Catalog. Consequences:

- Cruise IDs, once published, cannot be deleted (but can be deprecated).
- Mappings of eg. researchers to ORCIDs must be verified.
- R2R services should be online 24/7 and reasonably performant.
 - Linked Data collection
 - SPARQL (RDF Query Language) endpoint
 - OGC CSW (Catalog Service) endpoint



Rolling Deck to Repository (R2R) Linked Open Data
Running at <http://data.rvdata.us/>

Home | [Award](#) | [Campaign](#) | [Cruise](#) | [CruiseType](#) | [Device](#) | [DeviceType](#) | [Document](#) | [DocumentType](#) | [Event](#) | [Fileset](#) | [Format](#) | [Geometry](#) | [Model](#) | [Organization](#) | [Participant](#) | [Person](#) | [Port](#) | [Product](#) | [Program](#) | [Track](#) | [Vessel](#)

This knowledge base is published with D2R Server. It can be accessed from

1. Any old web browser,
2. Semantic Web browsers, and
3. SPARQL clients.

1. HTML View
You can use the navigation links at the top of this page to explore the knowledge base.

2. RDF View
You can also explore this knowledge base with Semantic Web browsers like [Marbles](#).
To start browsing, open this entry point URL in your Semantic Web browser:
<http://data.rvdata.us/all>

3. SPARQL Endpoint
Clients can query the knowledge base at this SPARQL endpoint:
<http://data.rvdata.us/sparql>
The endpoint can be explored in a [Snoql](#) interface from any Web browser.

Powered by [PostgreSQL](#), [Tomcat](#), and [D2R Server](#)

<http://data.rvdata.us/>



#9. Improve Iteratively

R2R started by characterizing every Instrument with minimum required metadata

- Device Type
- Parent Vessel
- Make
- Model
- File Format

... then circling back selectively over time to add

- Child Devices (subcomponents of complex packages)
 - Location (on Vessel)
 - Type-specific details (eg. Multibeam array configuration)
- etc.



#10. Metrics

R2R develops statistics to demonstrate impact.

Input side

- # of cruises cataloged
- # of datasets preserved and disseminated
- # of QA reports published
- # of products created

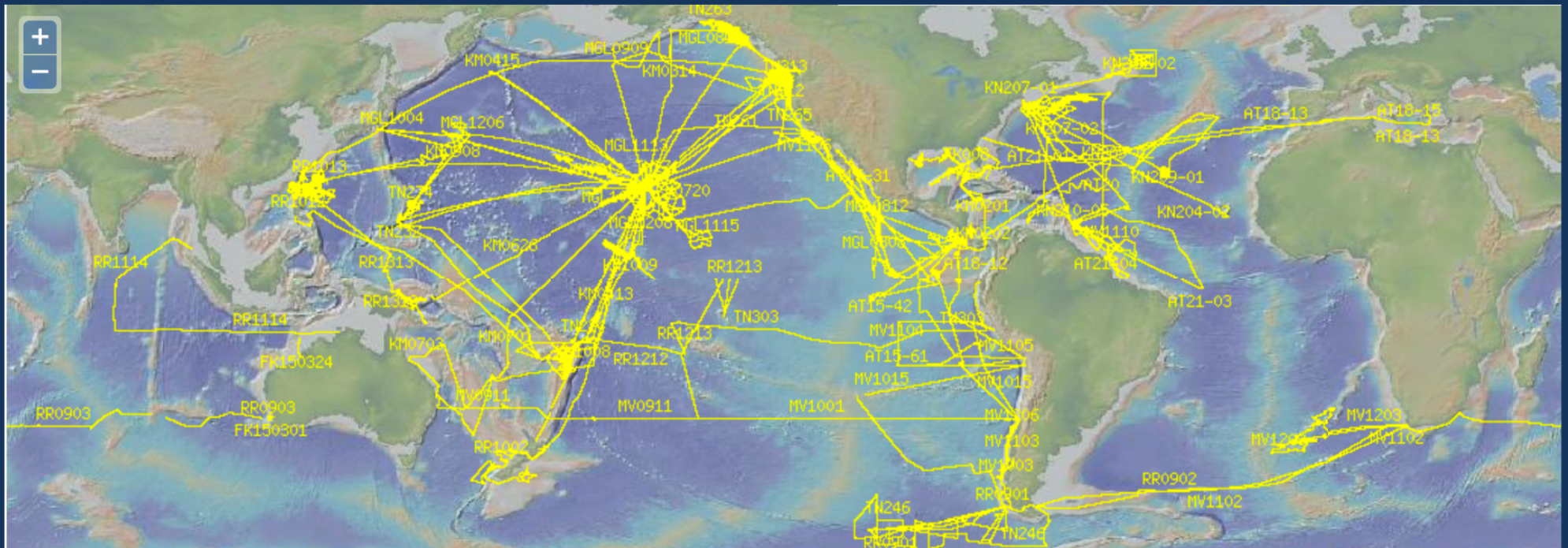
Output side

- # of datasets+products downloaded
- # of website visits
- # of datasets contributed to global syntheses *
- # of links to related resources *
- Derived datasets
- Samples
- Journal articles



#10. Metrics (cont.)

R2R contributes to
Global Multi-Resolution Topography (GMRT) Synthesis



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

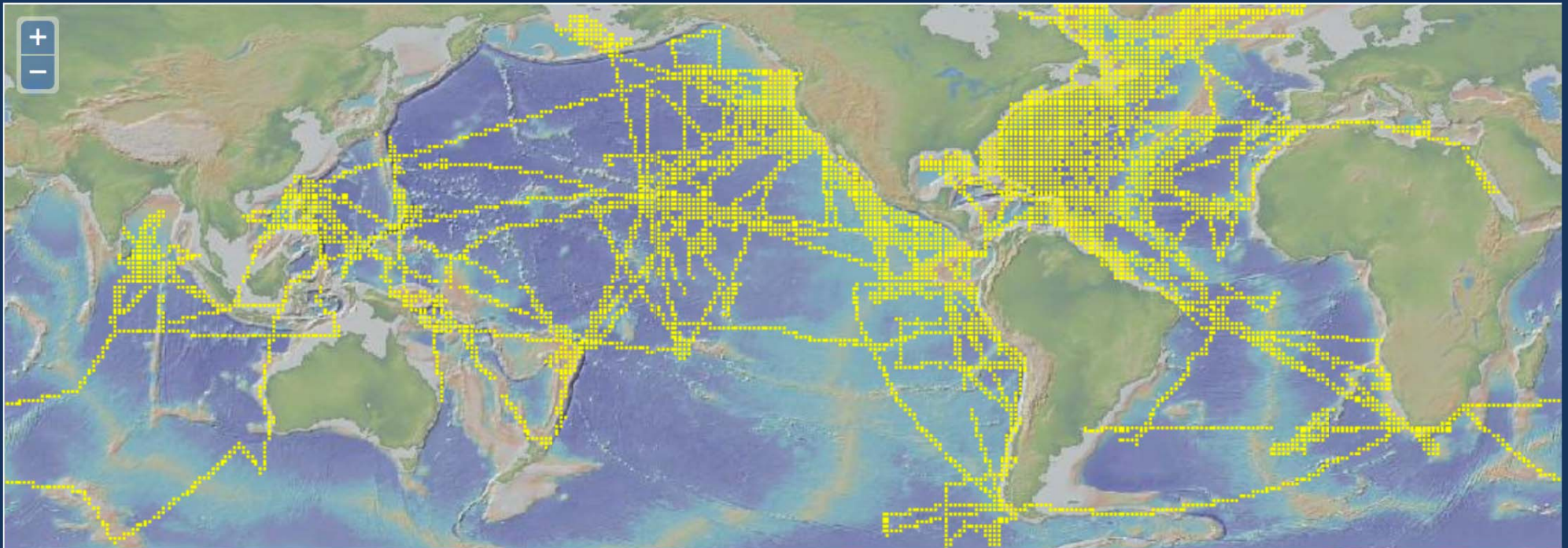


*GMRT produced by the
Interdisciplinary
Earth Data Alliance
(IEDA)*



#10. Metrics (cont.)

R2R contributes to
International Comprehensive Ocean-Atmosphere Data Set (ICOADS)



*Above: METOC data contributed by SAMOS
in partnership with R2R (as of ICOADS v 3.0).*

UNOLS Ships: 10
SAMOS Records: 278,1014
Ship Days: 11,792



*Shipboard Automated
Meteorological and
Oceanographic System
(SAMOS)*



#11. Open Data

Cruises on US research vessels are funded by many different agencies, with different data release policies.

R2R original policy assumed default 2-year proprietary hold on data, following NSF OCE policy.

- New NSF policy in 2011 changed to “Underway data should be released in 60 days unless PI requests a hold”.

New R2R policy is to seek immediate release on every cruise, but with explicit confirmation from Chief Sci.



#12. Licenses

R2R original policy asserted CC-BY-NC-SA (3.0) license on every Dataset that requires Attribution, Non-Commercial, and Sharealike.

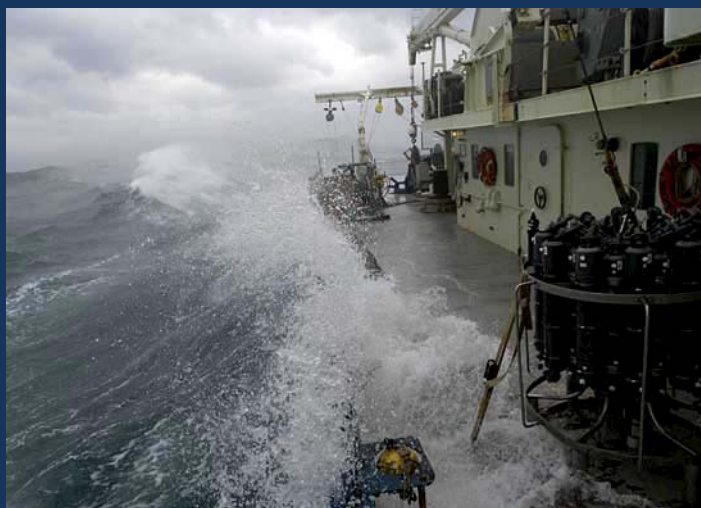
New policy changes to CC0 and will include statement with every Dataset that provides:

- Public Domain (CC0) statement
- DOI
- “Please cite as”



Thank you.





info@rvdata.us

