

Member Node Description: Rolling Deck to Repository (R2R)

Version 1.0

4/8/16

Bob Arko

General

Name of resource:	Rolling Deck to Repository (R2R)
URL(s):	http://www.rvdata.us/
Institutional affiliation(s):	Lamont-Doherty Earth Observatory, Florida State University, Scripps Institution of Oceanography, Woods Hole Oceanographic Institution; in partnership with the University-National Oceanographic Laboratory System (UNOLS)
Primary geographic location:	New York, New York, USA
Project Director & contact info:	Suzanne Carbotte, LDEO
Technical Contact & contact info:	Robert Arko, LDEO, admin@rvdata.us
Age of resource:	Since 2009
Funding support:	National Science Foundation (primary), Office of Naval Research, National Oceanic and Atmospheric Administration, Schmidt Ocean Institute
Proposed Unique Identifier:	urn:node:R2R

Content

Content description/collection policy (1 paragraph, domain and spatial/temporal coverage, uniqueness of content, exclusions, as applicable):

The Rolling Deck to Repository (R2R) Program provides a global catalog of field expeditions on US oceanographic research vessels; ensures long-term preservation and access to the original environmental sensor data routinely acquired during each expedition; creates a suite of processed data products and quality assessment reports for each expedition; provides support for at-sea event logging; and maintains links to additional datasets and publications at other repositories for each expedition.

Types of data (complex objects, text, image, video, audio, other):

Original field data from meteorological, oceanographic, and geophysical sensors as well as navigational and engineering instruments are preserved by R2R in the original instrument-specific formats. Processed data products created by R2R are published in standard, documented ascii tabular formats.

Data and metadata availability (rights, licensing, restrictions):

All data and documentation published by R2R are freely available in the public domain.

Option for embargo (yes/no, duration):

The original field data delivered from the research vessel to R2R after each expedition may be embargoed on request from the chief scientist, in accordance with the policy of the funding agency. R2R does not publish a Digital Object Identifier (DOI) for a dataset, or submit it to a public long-term archive, until such embargo has expired.

Size of holdings (number and size of datasets, mean and median granules (files) per dataset):

R2R has curated approximately 50TB of data and documentation from over 6,000 field expeditions, as of 2016. New data and documentation are being added a rate of approximately 7TB/year from 350 expeditions/year.

Please describe recent usage statistics, if known, including information on annual data product downloads, annual number of users, annual number of data products used in publications:

N/A

User interactions

How does a user contribute data? (what can be deposited, how are data prepared, are specific software required, documentation/support available)

Users do not contribute data to R2R. Rather, data are delivered to R2R directly from a research vessel at the end of an expedition. Users may contribute field documents (cruise reports, sample logs, equipment manuals, etc.) to R2R at the end of an expedition.

How does a user acquire / access data?

A user may browse the Cruise Catalog on the R2R web site to find expeditions and download datasets. Users may also find and download datasets in long-term archives such as the NOAA National Centers for Environmental Information (NCEI), using one of the NCEI search portals.

What user support services are available (both for depositing and accessing/using data)?

R2R staff members respond to questions and requests via the R2R web site contact form and/or email (info@rvdata.us).

How does the resource curate data at the time of deposit?

Data and documentation are initially held on R2R local disk storage when received from a research vessel, and ultimately deposited with an appropriate long-term archive (either NCEI or Chronopolis, depending on the data type). R2R creates standard documentation for every dataset that identifies the research platform, sensor system, parent expedition, field investigators, file format, etc. A selected subset of datasets from common sensors is assessed for quality, using semi-automated (scripted) routines developed in consultation with disciplinary specialists.

Technical characteristics and policies

Software platform description, incl. data search and access API(s):

The R2R Catalog is published on the Web as Linked Open Data, and can be queried via a SPARQL endpoint. Trackline geometries for the field expeditions are published as Web Feature and Map Services (WFS,WMS).

Service reliability (including recent uptime statistics, frequency of hardware refresh, if known):

N/A

Preservation reliability (including replication/backup, integrity checks, format migration, disaster planning):

Content held on R2R local disk storage is replicated in two locations on the LDEO campus, in addition to an offsite copy in Amazon Glacier.

User authentication technology (incl. level of create/modify/delete access by users):

Users do not require a login to access the R2R Catalog or any associated services, and do not have the ability to modify any content.

Data identifier system and data citation policy, if available:

R2R publishes a Digital Object Identifier (DOI) for each field expedition after it concludes; for each original field dataset after any embargoes are cleared; for each field document (eg. cruise report) submitted to R2R by a member of the science party; and for each standard processed data product that R2R creates.

Metadata standards (including provenance):

R2R publishes DOIs against the DataCite Metadata Kernel (currently version 3) as well as geospatial metadata records against the ISO 19115-2 standard, and uses controlled vocabularies for platform types, sensor types, parameter types, etc., published by the NERC Vocabulary Server.

Capacity/services to DataONE

At what functional tier will you initially be operating? (see <http://bit.ly/MNFactSheet> for definitions)

- Tier 1: Read only, public content
- Tier 2: Read only with access control
- Tier 3: Read/write using client tools
- Tier 4: Able to operate as a replication target

If you can host data from other member nodes, what storage capacity is available?

N/A

Can you provide computing capacity to the broader network? If so, please describe.

N/A

Other Services

What other services or resources (such as expertise, software development capacity, educational/training resources, or software tools) can be provided of benefit to the broader network?

N/A