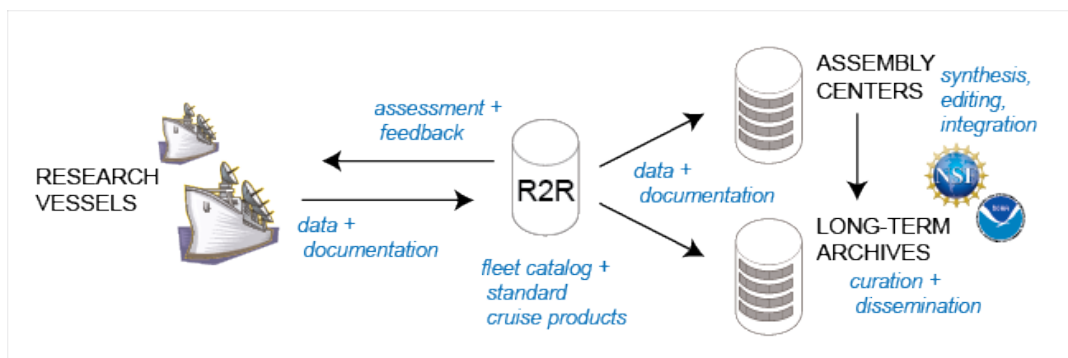


The Rolling Deck to Repository Program Marine Data Services for the US Academic Research Fleet

Mission

The Rolling Deck to Repository (R2R; www.rvdata.us) program documents and preserves environmental sensor data acquired during scientific expeditions on U.S. academic research vessels. The program is a collaborative effort between Lamont-Doherty Earth Observatory, Florida State University, Scripps Institution of Oceanography, and Woods Hole Oceanographic Institution; and works closely with the University-National Oceanographic Laboratory System (UNOLS; www.unols.org), whose membership includes 58 U.S. academic institutions supporting oceanographic research worldwide. R2R is funded primarily by the National Science Foundation with support from the Office of Naval Research and the Schmidt Ocean Institute.



Products and Services

Each research vessel delivers a package of original/unprocessed navigational, geophysical, oceanographic, and meteorological data from its permanently installed sensor systems to R2R at the end of an expedition, along with a manifest that provides the vessel ID, cruise ID, title, start/end dates and ports, and science party members. R2R deposits a copy of each package to private offline storage segments at both the NOAA National Centers for Environmental Information (NCEI) and Amazon Glacier. Each package is then broken out into individual datasets according to sensor type, make, model, and file format; and a master catalog of expeditions and datasets is published online via the R2R Web site. A Digital Object Identifier (DOI) is published for each expedition as well as for each dataset. After permission is granted by the Chief Scientist, individual datasets are posted for public download via the R2R Web site. Selected datasets are submitted to NCEI for dissemination and inclusion in global syntheses.

R2R assesses the quality of selected data types, using a scripted workflow and criteria developed in collaboration with specialists in the science community. The assessment results are published online via the R2R Web site, and standard ratings are calculated as part of feedback to vessel technicians. R2R also produces a standard set of data products after each expedition including quality-controlled shiptrack navigation, underway geophysical profiles (gravity, magnetics, bathymetry), water column depth profiles from CTD hydrocasts, and real-time

meteorology/near-surface oceanography; all of which are posted for public download via the R2R Web site. R2R supports an Event Logger application, including a shipboard microserver, to assist science parties in documenting their scientific sampling while underway.

The R2R program interoperates with 14 other data repositories, primarily NSF-sponsored, that manage other kinds of marine data content related to cruises inventoried in the R2R Catalog. The data content hosted in these repositories includes data acquired with specialized science party instruments and national instrument facilities, scientific sampling logs and associated laboratory analyses, as well as processed data products derived from field data, global synthesis products, and links to articles in scientific journals. A suite of Web-based services support interoperability including a OGC Web Feature Service (WFS) that provides shiptrack geometries; a Catalog Service for Web (CSW) that provides ISO 19139 XML records for expeditions; a W3C "Linked Data" graph and associated RDF Query Language (SPARQL) endpoint for Semantic Web clients; and customized Atom+GeoRSS feeds for partner programs such as OOI and ECS. The complete inventory of expeditions and datasets in the R2R Catalog are discoverable in global research indexes such as DataCite (<http://search.datacite.org>).

Infrastructure

R2R's computer infrastructure is primarily located on the LDEO campus of Columbia University in Palisades, New York, with selective extensions to commercial providers. The LDEO campus cluster consists of six Dell Linux-based servers, six ACNC fiber storage arrays, and a supporting local network of switches/routers, firewalls, and environmental monitors, split between two buildings. Hardware is typically refreshed on a 5-year cycle. Monitoring and backups are implemented via the Nagios and Bacula open-source packages. The application infrastructure is open-source software consisting of Apache Httpd/Tomcat, PostgreSQL, and PostGIS backends. Programming is primarily PHP and Shell scripting, managed in GitHub private repositories, using open-source libraries such as GDAL and MB-System. Commercial provisioning is used for outward-facing Web services such as the R2R Search page (Linode.com), and an off-site backup copies of R2R data packages received from all research vessels are stored in a Amazon Web Services Glacier vault in the US-West-2 (Oregon) zone.