

From Rolling Deck to Repository (R2R): Progress in Systematically Capturing Underway Data from the US Academic Fleet, including Multibeam



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

Suzanne Carbotte², Stephen Miller¹, Andrew Maffei³, Shawn Smith⁴, Vicki Ferrini², Robert Arko², Karen Stocks², Suzanne O'Hara², Cynthia Chandler³, Mark Bourassa⁴, Dru Clark¹, Aaron Sweeney¹, John Morton²

rvdata.us

- ¹ Scripps Institution of Oceanography
- ² Lamont-Doherty Earth Observatory (Lead Institution)
- ³ Woods Hole Oceanographic Institution
- ⁴ Florida State University
- ⁵ San Diego Supercomputer Center

Launched in 2009, R2R is a systematic effort to capture, catalog and archive US underway shipboard data. As of September 2011, data from 2,130 cruises on 26 vessels had been submitted, totaling 7,481,290 files (>9 TB).

Each vessel in the US academic fleet is equipped with a multidisciplinary suite of sensors that are available for continuous operation during each expedition. The resulting "underway" geophysical, water column, and meteorological datasets describe basic environmental conditions for the oceans and are of high value for building global syntheses, climatologies, satellite validation and historical time series of ocean properties.

The R2R Portal (www.rvdata.us) will be the central gateway through which underway data are routinely cataloged and securely transmitted to the appropriate national data center, ensuring long-term access and relieving chief scientists of their individual obligations under NSF policy to submit underway data.

Protocols are being developed for quality assessing high priority underway data types, to provide feedback to shipboard instrument operators and to inform end users. Standard metadata will be supplied with each dataset, including provenance and quality information. Standard products, such as quality-controlled navigation, are being created.

As part of this work, R2R has collaborated with NOAA to create an XML-based, ISO 19115-compliant cruise metadata template. This describes the basic elements of a seagoing expedition: cruise identifier, vessel name, operating institution, dates/ports, navigation track, survey targets, science party, funding sources, scientific instruments, daughter platforms, and data sets. Controlled vocabulary terms are directly embedded as Uniform Resource Identifier (URI) references. We envision a hierarchical framework where a single "cruise-level" record is linked to multiple "dataset-level" records that may be published independently.

One of the subprojects within R2R is the development of a shipboard scientific event logging system that incorporates best practice guidelines, controlled vocabularies, a cruise metadata schema, and a scientific event log. The ELOG-based cruise event logging system, currently being tested, enables researchers to record digitally all scientific events and assign a unique event identifier to each entry, to assist in the ingestion of these data into oceanographic data repositories and subsequent reuse of the datasets.

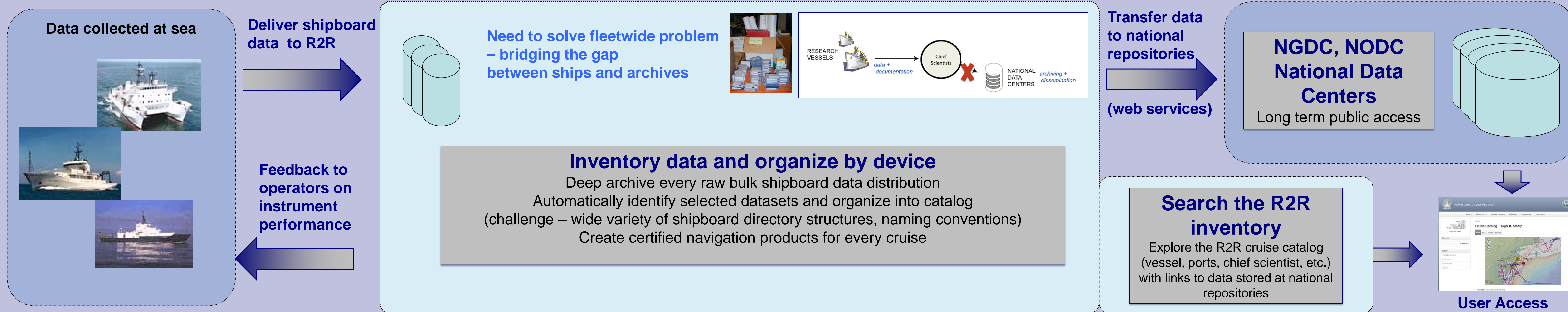
Rolling Deck to Repository is a collaboration between Lamont-Doherty Earth Observatory (lead institution), Scripps Institution of Oceanography, San Diego Supercomputer Center, Woods Hole Oceanographic Institution, and Florida State University; and works with the vessel operating institutions, UNOLS Office, NOAA National Data Centers, and disciplinary data assembly centers (DACs). Support is provided from support from the National Science Foundation (NSF), Oceanographic Instrumentation and Technical Services (OITS) Program, NSF OCE-0947828.



A Community Effort
LDEO/SIO/WHOI/Florida State
UNOLS
NOAA



Systematic effort to capture and archive all US academic routine underway data



Explore R2R Cruise Catalog and link to download files from national repositories

Select a vessel, examine inventory

R2R catalog provides links to NGDC or NODC for long term public access

Event Logger – a flexible, extensible approach to capturing field sampling events based on open-source software.

CHDS11-01
RV Cape Hatteras Science Event Log: LAMMO: June 1, 2011 – June 21, 2011; Chief Scientist: Jim Ledwell (WHOI); Page 1 of 13

Event	Date/Time/UTC	Instrument	Action	Station	Latitude	Longitude	Author	Comment	Revisions
20110601.1316.001	20110601.1317	Ship	startCruise	NaN	34.694000	-76.670400	J.Ledwell		
20110601.1742.001	20110601.1840	ADCP150	other	NaN	34.114217	-76.150767	J.Ledwell	turning 90 degrees	
20110601.1820.001	20110601.1821	ADCP150	other	NaN	34.112433	-76.170750	J.Ledwell	Pol: disabled; setting up software; stopped here about 20 min	
20110601.1825.001	20110601.1826	ADCP150	startLine	NaN	34.129867	-76.170650	J.Ledwell	Calibration run: Decr ~ 100 ft	6/01 Jun 2011 18:26
20110601.1852.001	20110601.1853	ADCP150	endLine	NaN	34.096667	-76.164300	J.Ledwell	End of calibration run. Locks ok	
20110601.2302.001	20110601.2305	calibration/fluorometer	start	NaN	33.678667	-75.631067	J.Ledwell	Started setting up around 2200 Z	
20110602.1355.001	20110602.1356	ADCP150	startLine	NaN	32.006933	-73.629817	SPierce	ok at 6 kts	6/02 Jun 2011 14:05

The R2R Event Logger uses the ELOG weblog software (<https://midas.psi.ch/elog/>). A custom configuration is generated for each research cruise. This web application is accessible to researchers using any browser on the ship network.

A final event log that documents all instrument-oriented sampling events, with the use of controlled vocabularies, is included in and archived with the final cruise data distribution. The event logger is currently being tested on cruises.

Initial accomplishments

Community collaboration

- Developing dataflow and submission agreements with NGDC and NODC
- Site visits and technical discussions with UNOLS, NOAA and 18 vessel operating institutions
- Defining ISO-standard cruise- and dataset-level metadata, and associated controlled vocabularies, in cooperation with NOAA. Develop templates for additional supporting documentation (see panel on right)
- Coordinating plans with international programs: SeaDataNet, Eurofleets, Geo-Seas

Developing the processing pipeline

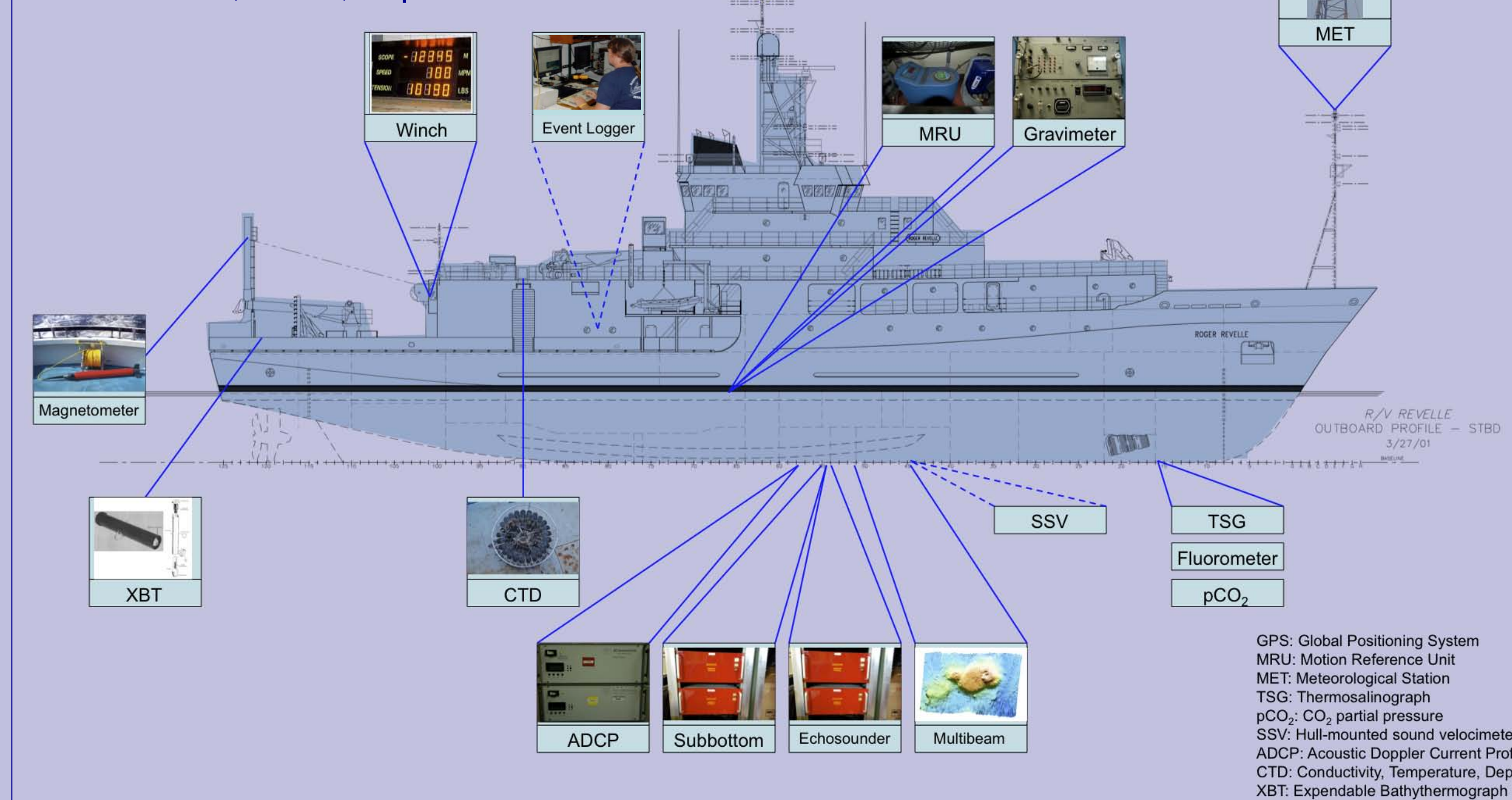
- Developing dataflow procedures for 26 vessels
- Capturing real-time data with the Shipboard Automated Meteorological and Oceanographic System (SAMOS)
- Creating automatic cataloging scripts (sort through complex bulk shipboard distributions, identify data sets, populate standard directory structure, populate metadata database)
- Developing automated quality assessment for selected datasets

Routine provision of data to National Data Centers

- Saving all original, raw bulk shipboard distributions in deep archive at NGDC (9 TB), a first step while data breakout and quality processes are in progress
- Routine creation and transfer of cruise-level metadata and standard navigation products
- Maintaining vessel instrument profile database, tracking changes, in coordination with UNOLS

R2R & UNOLS manage vessel underway data profiles

Operators provide instrument device type, manufacturer, model, shipboard location



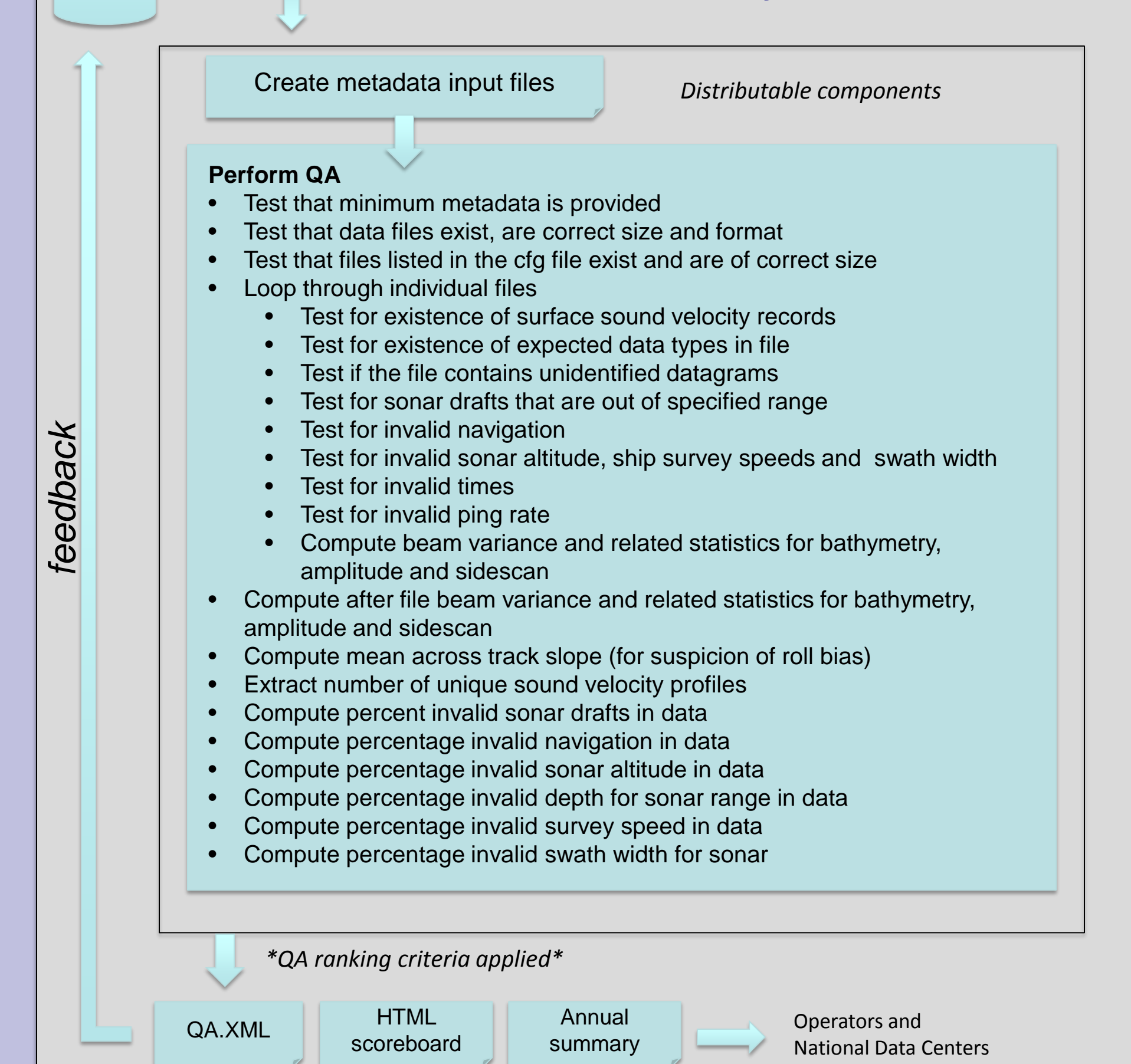
Data files are not enough

Metadata

To ensure products delivered to the National Data Centers are well described and usable for decades to come, R2R includes the following supporting information with all fileset submissions:

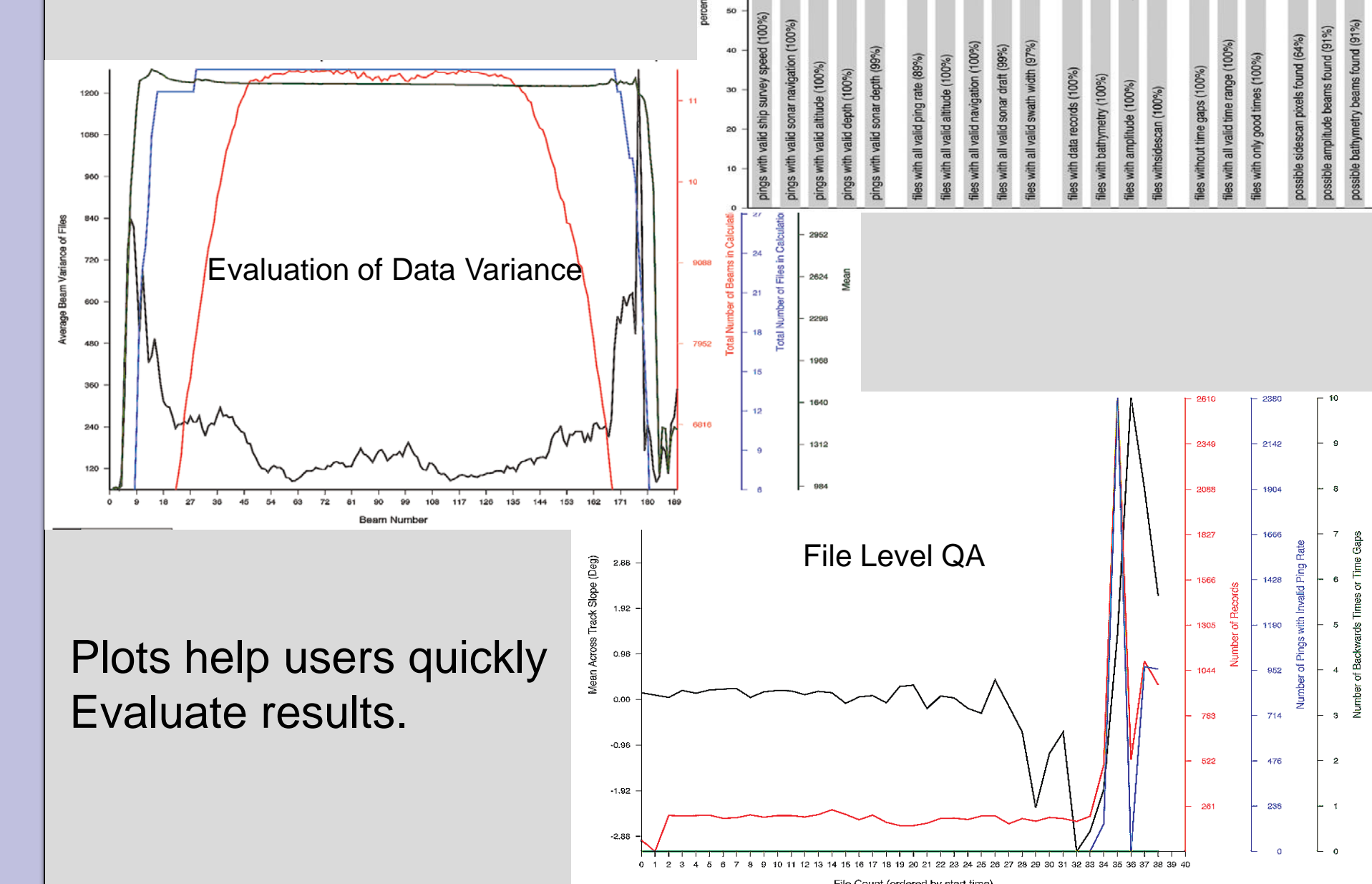
- ISO-compliant cruise-level metadata
- File manifest
- File format description document (human-readable)
- Fileset-level metadata, including quality assessment

Multibeam Quality Assessment



Quality Assessment is performed on the Multibeam data provided by the ship operators. Automated scripts flag suspicious parameters.

An XML format log travels with the data.



R2R has adopted a Linked Data approach (<http://www.w3.org/standards/semanticweb/data>) to publishing catalog content, based on the W3C Resource Description Framework (RDF) and Uniform Resource Identifiers (URIs).

Our data model will be published as a collection of W3C Simple Knowledge Organization System (SKOS) concepts, mapped to partner vocabularies such as those developed by the Global Change Master Directory (GCMD) and the pan-European SeaDataNet partnership.

Our catalog content will be published as collections of RDF resources with globally unique and persistent identifiers.