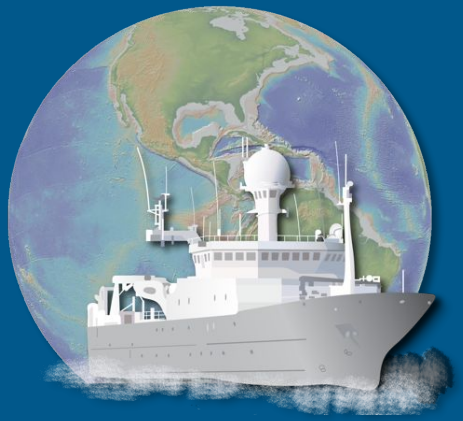
Rolling Deck to Repository (R2R)

From Operator to R2R: Improving the Flow Together

S. O'Hara, D. Clark, C. Olson,, J. Elya, K. McLain, C. Sellers, R. Hudak, S. Smith, K. Stocks, L. Stolp, S. Carbotte
LDEO, FSU, SIO, WHOI

From Operator to R2R: Improving the Flow Together

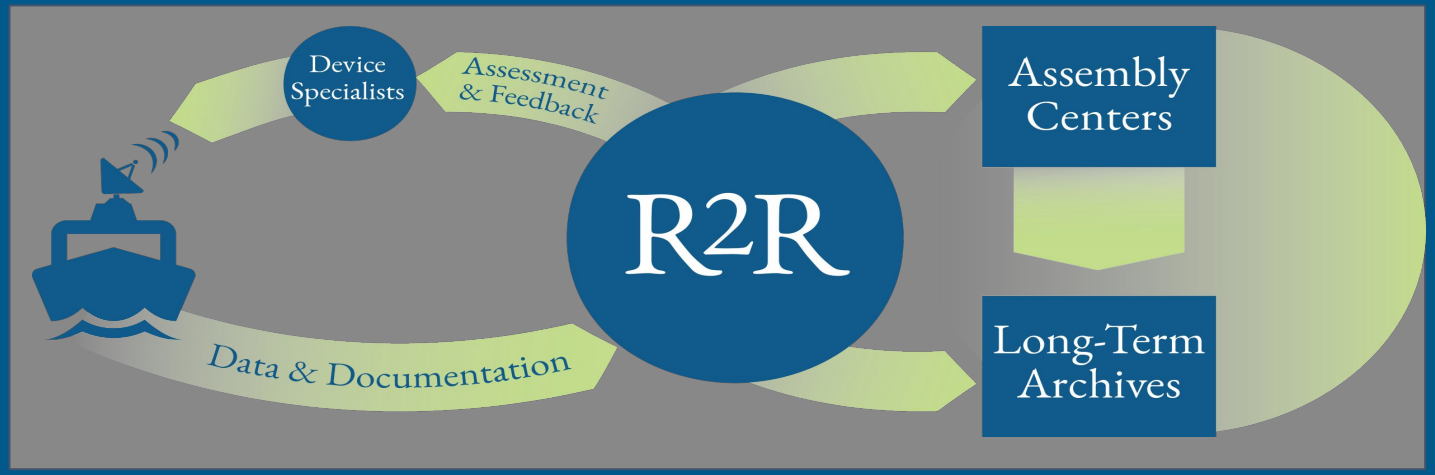
- Introduction
- Flow of data and cruise metadata
- Keeping device inventories up-to-date
- Navigation best practices
- New R2R tools to help operators & final comments



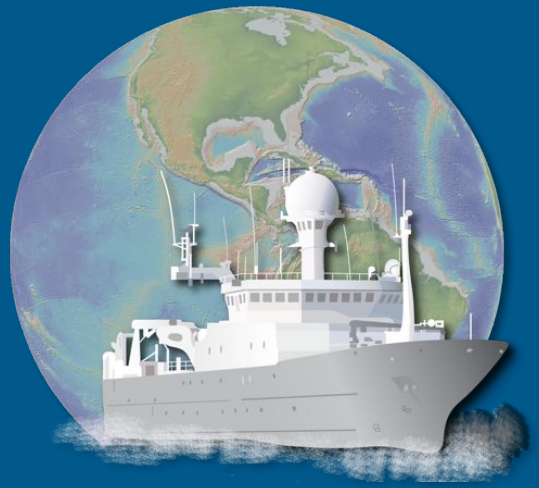
Each section will start with an introduction by a R2R team member, followed by open discussion.

We encourage all to participate.
Share your successes and failures to help others.

R2R's Mission



R2R gets your data documented, archived and available to the public.



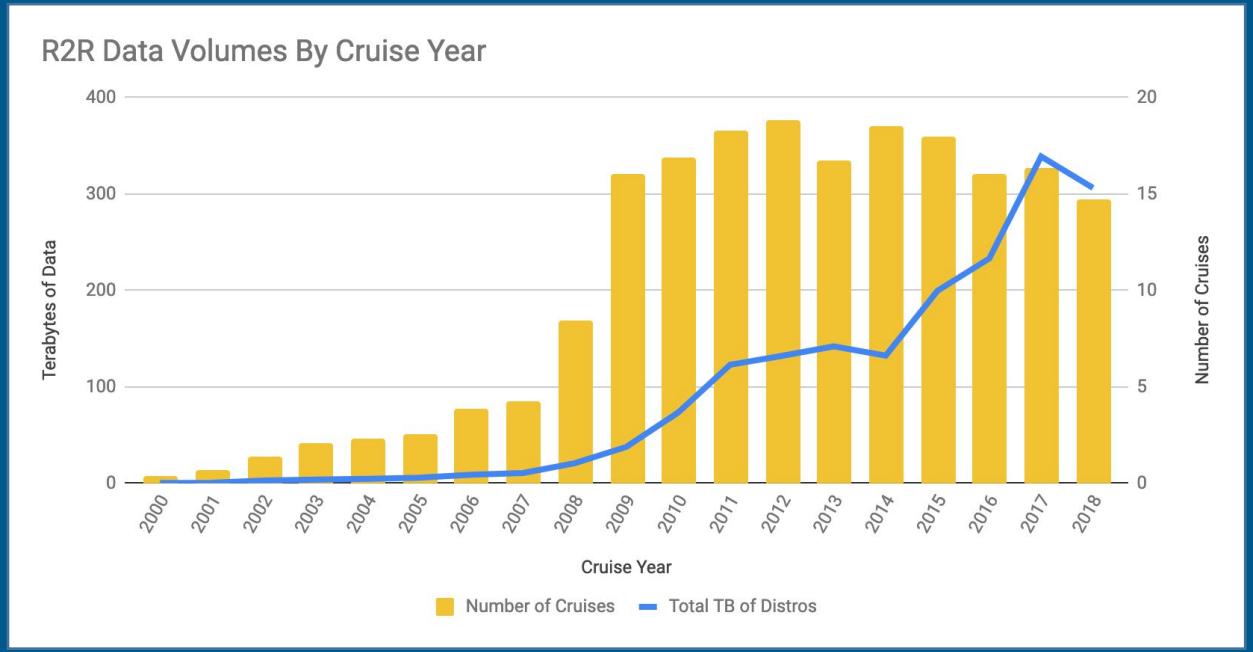
From Operator to R2R: Improving the Flow Together

Flow of Data and Cruise Metadata

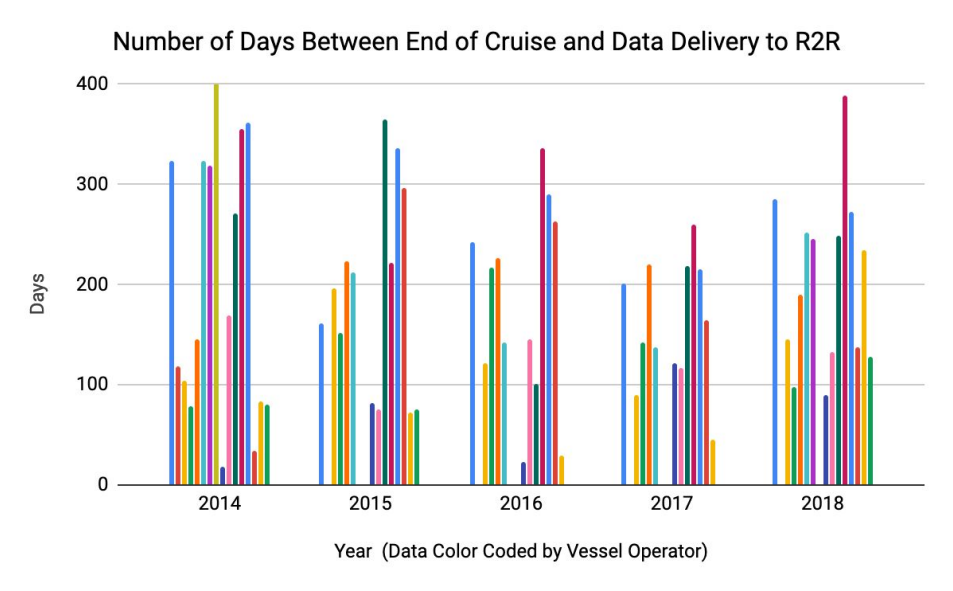
Suzanne O'Hara (sohara@ldeo.columbia.edu)
LDEO

R2R - Cruises & Data Volumes

Number of cruises vary by year, but totals have increased significantly over time.



Cruise Data Arriving at R2R



R2R will process, archive to NCEI and distribute your data to the public. Don't leave your data sitting on the ship.

We can help you send data and metadata faster and easier. Just ask.

Don't delay sending cruise data until the end of the year!

What is Cruise Metadata?

Minimum metadata includes:

- Unique cruise identifier
- Vessel Name
- Depart/arrive dates and ports
- Chief Scientist name and institution

Full metadata includes:

- Descriptive cruise title
- Full personnel list with roles and institutions

Minimum metadata allows the cruise to be generated at R2R, but more is always best!

Send cruise metadata using the UNOLS form
https://www.unols.org/document/cruise-personnel-manifest
or contact us for other options

The Cruise Identifier is Really Important!

A unique and permanent cruise identifier allows the organization, tracking and cross referencing of all data and metadata for a cruise.

Creating a unique cruise identifier at the start of a cruise and never changing it is ideal.

R2R cannot start any work on your cruise until an identifier is assigned. In many cases waiting for a cruise identifier to be assigned is the main delay to getting your data archived and online!

Cruise Data Distros Best Practices

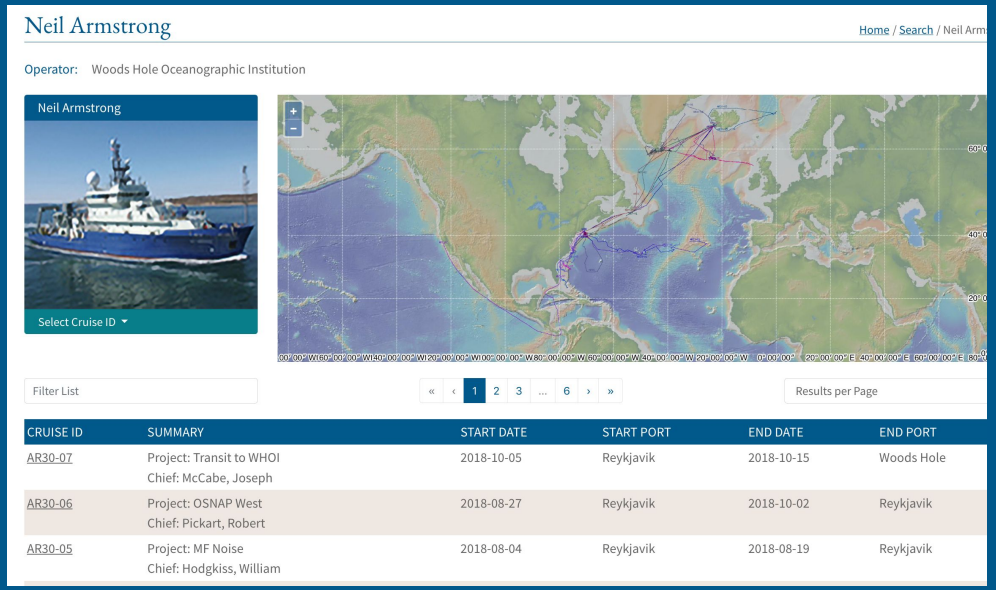
- Don't wait until the end of the year to send data. R2R will work with you to find the best way to transmit your data promptly (ftp, rsync, usb disk, ??).
- Be consistent with how data is organized and named. Notify R2R of any changes.
- Do not use special characters or spaces in directory or file names
- Provide a checksum with your cruise distro to allow R2R to verify distro file

Visit https://www.rvdata.us/community/ship-operators to find suggested directory structure and other best practices

Your Cruises at R2R and Onward

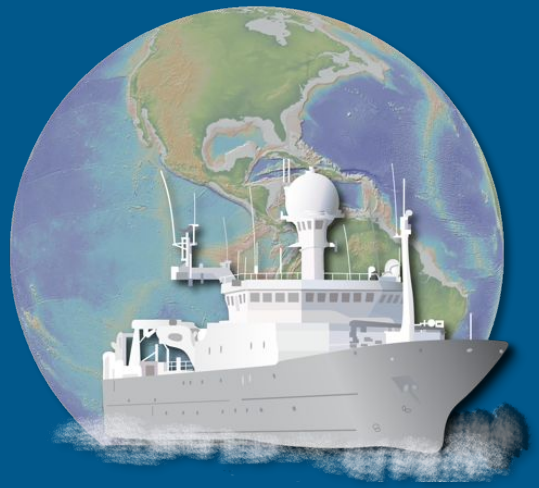
Providing complete cruise metadata and data to R2R promptly is the best way to get your cruise data organized, archived and online.

Your underway cruise data sets that may not have been of interest to the initial science party can be of great value to other investigators once it is available online.



Flow of Data and Cruise Metadata

- What would improve the flow of data and metadata to R2R?
- What do you see as the biggest roadblocks to transmitting data and metadata?
- What additional tools or services would be most useful?
- ?



From Operator to R2R: Improving the Flow Together

Keeping Device Inventories Up-to-Date

Dru Clark (dclark@ucsd.edu)
SIO

Outline

Data flow: Quick review of the inventory process

Device info: Vessel Profile and tools

Planned Development

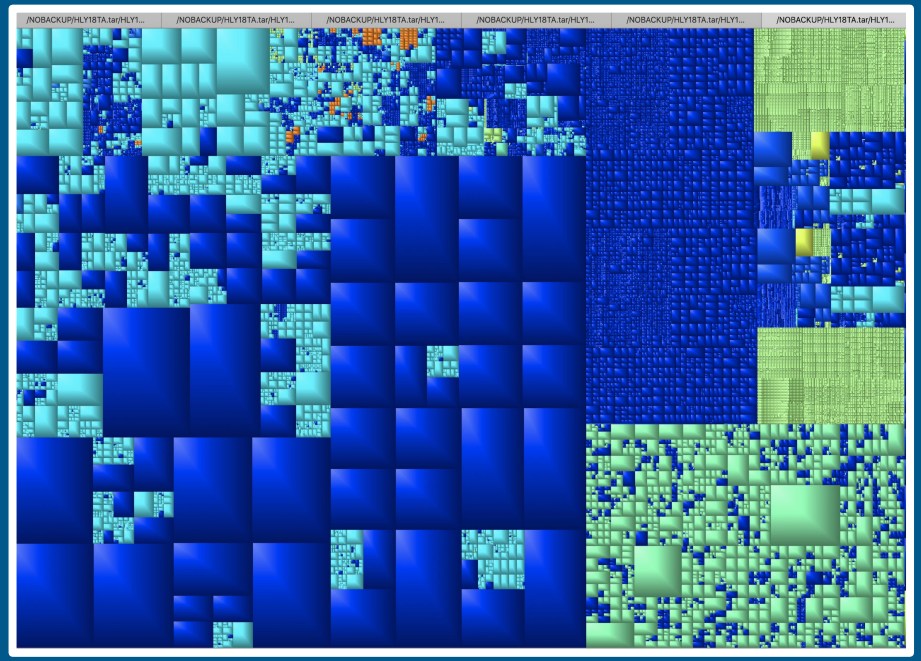
Rolling Deck

Cruise data for each underway device is stored and transferred to R2R

To Repository

R2R receives the official set of data for a cruise

Cruise data distribution (distro)



Complex

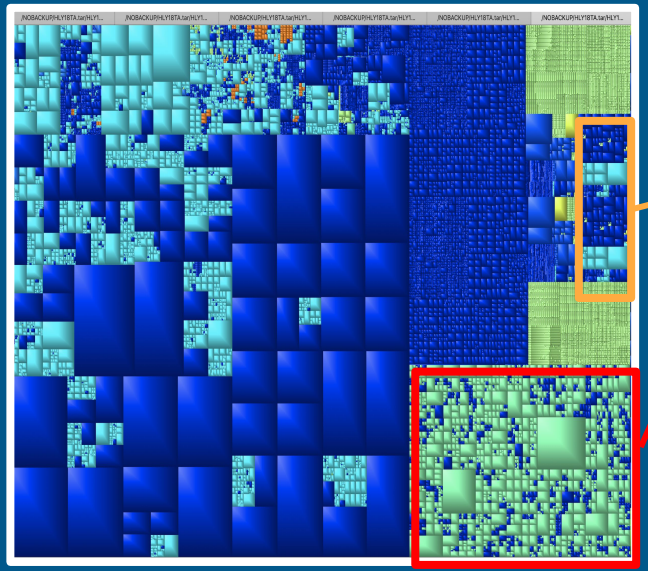
Heat map of a single cruise data distribution (distro) colored by file type
Blues are data (binary) Light green/yellow includes pdf, txt and other ascii files.

Distro workflow:

- Inventory files
- Breakout by device

R2R to NOAA

Cruise distro



Associate files to device

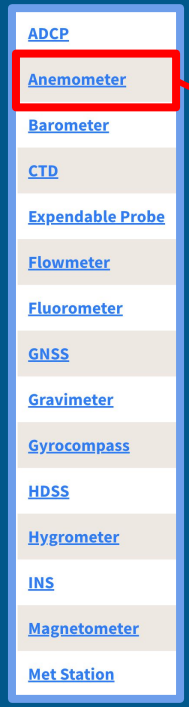
Database:
Device mapping

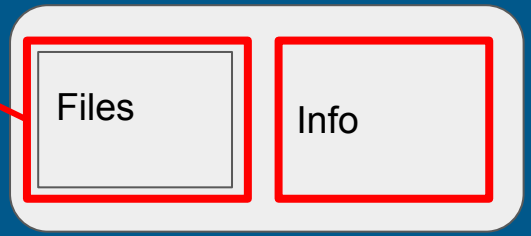
- ADCP
- Anemometer
- Barometer
- CTD
- Expendable Probe
- Flowmeter
- Fluorometer
- GNSS
- Gravimeter
- Gyrocompass
- HDSS
- Hygrometer
- INS
- Magnetometer
- Met Station

NOAA
National Centers for Environmental Information

Archive at NOAA

Metadata is required





An accurate file description required for submission

- Device
- Format
- Cruise

Vessel device profile

snapshot of the device configuration

R2R ID	Device Type	Make/Model	Time source	Fileset count	Tools
100474	acqsys	Hawaii UHDAS		158	Edit
100542	adcp parent device:100474	- RDI VM-300		.	Edit
100497	ctd	Sea-Bird SBE-911plus		95	Edit
100299	expendableprobe	Sippican MK12			Edit
100820	gnss	Furuno GP-170		34	Edit
100673	gnss	Garmin GPS-17HVS		166	Edit
100301	gnss	Trimble Tasman			Edit
100300	gnss	Trimble NT-300D			Edit
100672	gnss	Ashtech ADU2		167	Edit
100475	gnss	Furuno GP-90D		115	Edit
100476	gnss Primary Navigation	Furuno GP-150		166	Edit
100674	gyrocompass	Furuno SC-30		147	Edit
100303	gyrocompass	Sperry MK-37 (2)			Edit
100302	gyrocompass	Sperry MK-37 (1)		17	Edit
100392	metstation	SIO MLT-System		188	Edit
100478	singlebeam	Knudsen 320B/R		63	Edit
100477	winch	Markey DESH-3		161	Edit

tech.rvdata.us

Red = review

Keeping the Vessel Profile updated

- We reach out
- RVTEC
- Ship visits when in local ports

This is person intensive for both techs and R2R
Overlaps with work by other groups

Planned Development

We would like to have a method to capture the device information at the time of the cruise that rides along with the cruise data and that can be reused by downstream processes.

i.e. Time stamped device listing in a known format

Next step

We are starting to pull together a working group

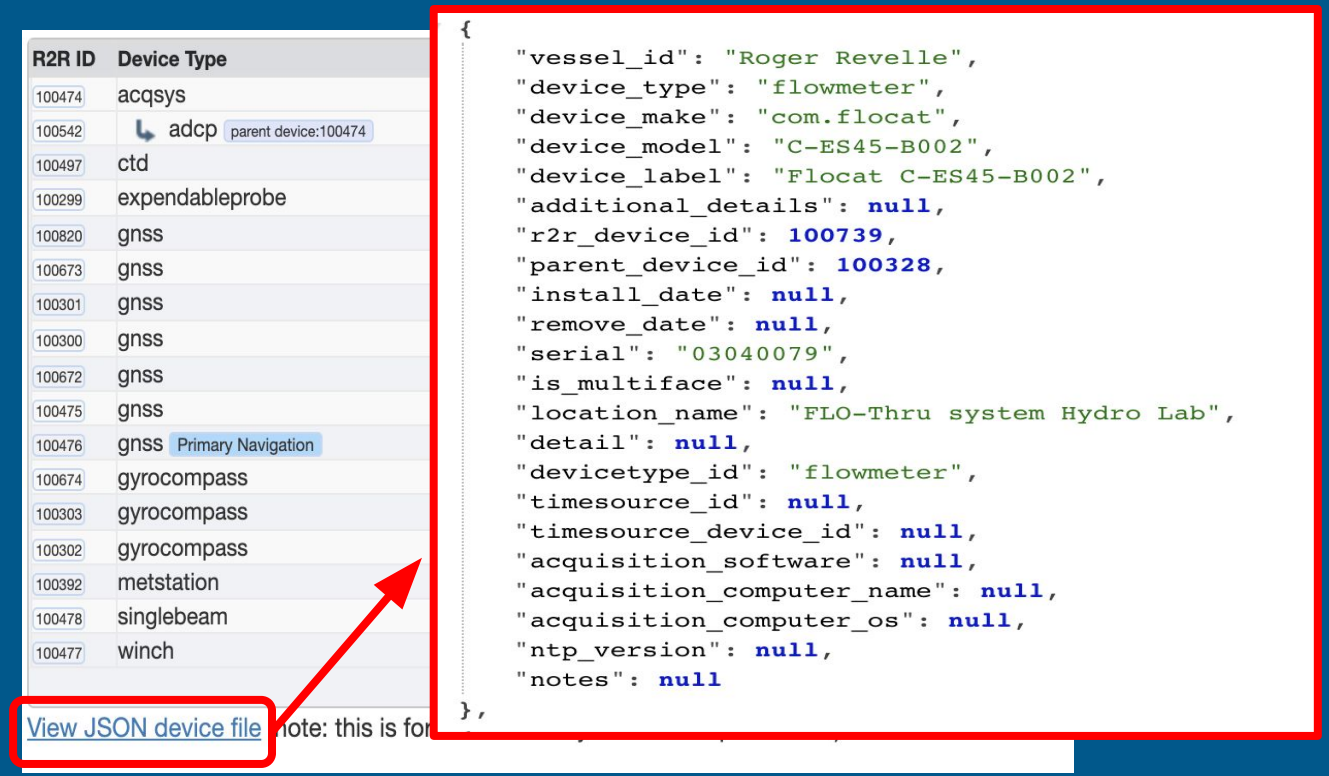
What: A device listing maintained on the vessel

Who: R2R, SAMOS, UNOLS (users)
and a set of ship techs (maintainers)

Proposed Device configuration file

- Initiate: Download at starter file from R2R
- Maintain and version: Git repo local on ship, push to hub
- Format: Json for simplicity - others?
- Validation: API or local to validate terms and structure

tech.rvdata.us



R2R ID	Device Type
100474	acqsys
100542	adcp parent device:100474
100497	ctd
100299	expendableprobe
100820	gnss
100673	gnss
100301	gnss
100300	gnss
100672	gnss
100475	gnss
100476	gnss Primary Navigation
100674	gyrocompass
100303	gyrocompass
100302	gyrocompass
100392	metstation
100478	singlebeam
100477	winch

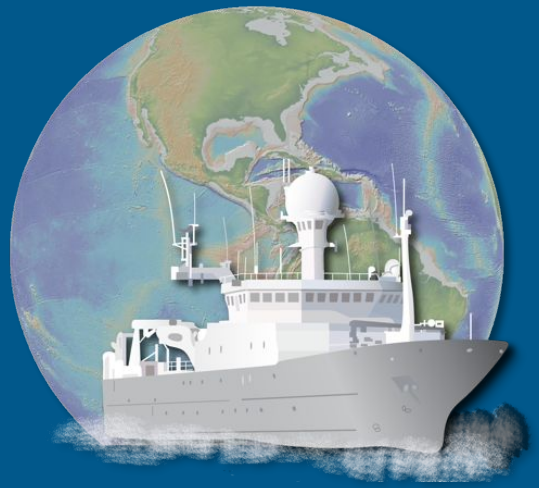
View JSON device file ote: this is fo


```
{
    "vessel_id": "Roger Revelle",
    "device_type": "flowmeter",
    "device_make": "com.flocat",
    "device_model": "C-ES45-B002",
    "device_label": "Flocat C-ES45-B002",
    "additional_details": null,
    "r2r_device_id": 100739,
    "parent_device_id": 100328,
    "install_date": null,
    "remove_date": null,
    "serial": "03040079",
    "is_multiface": null,
    "location_name": "FLO-Thru system Hydro Lab",
    "detail": null,
    "devicetype_id": "flowmeter",
    "timesource_id": null,
    "timesource_device_id": null,
    "acquisition_software": null,
    "acquisition_computer_name": null,
    "acquisition_computer_os": null,
    "ntp_version": null,
    "notes": null
},
```


Example JSON download of existed device list

Keeping Device Inventories Up-to-Date

Questions, Suggestions, Comments?



From Operator to R2R: Improving the Flow Together

Navigation Best Practices

Chris Olson (cjolson@ucsd.edu)
SIO

Why is navigation data important?

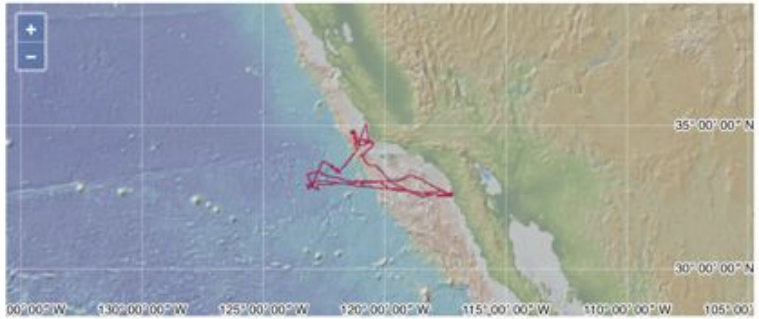
- Establishing spatial/temporal bounds
- Not all data have spatial information and in these cases needs to be matched to location based on timestamps

Sikuliaq

Home / Search / Sikuliaq / SKQ201605S

Operator: University of Alaska Fairbanks

Cruise DOI: 10.7284/906733



CRUISE ID	SUMMARY	START DATE	START PORT	END DATE	END PORT
SKQ201605S	Project: RAPID: Responses of the California Current Ecosystem to El Nino 2015-16 Chief: Ohman, Mark	2016-04-17	San Diego, California	2016-05-12	San Diego, California

R2R POST-FIELD PROCESSED PRODUCTS

DATA TYPE	DEVICE TYPES(S)	FORMAT	DATA
CTD	ctd (Sea-Bird SBE-911+)	seasoft-proc	
Gravity	gravimeter (Bell BGM-3)	r2rgrav_geocsv	
Navigation	ins (Kongsberg Seapath 320+)	r2rnav_geocsv	

Nav File Recommendations

- Standalone nav with date, time, location
- Data recorded at 1Hz or better
- NMEA strings
- Start new files daily

- Speed over ground and course over ground information optional

Primary Navigation Source

- We only look at one navigation source
- We maintain a list of “primary” navigation sources that are specified by the operator
- Can be viewed at tech.rvdata.us
- To update, email info@rvdata.us

Nav File Specifics

- Alternating GGA + ZDA strings
- DTM string at least once per file
- VTG strings recommended but not currently used by R2R
- Checksums optional but encouraged

A Note on Timestamps

- Preceding NMEA strings with a time stamp from a time server is a good idea and can help identify time issues

SP1824/raw/gp150-gps/gp150_20180901000000.raw

```
1535759999.856 $GPVTG,342.8,T,331.8,M,0.0,N,0.0,K*4A
1535759999.856 $GPZDA,000000,01,09,2018,07,00*4C
1535760000.866 $GPDTM,W84,,00.0000,N,00.0000,E,,W84*41
1535760000.956 $GPGGA,000001,3242.4592,N,11714.2041,W,2,8,2.3,12,M,,M,,*50
1535760000.956 $GPVTG,342.8,T,331.8,M,0.0,N,0.0,K*4A
1535760000.956 $GPZDA,000001,01,09,2018,07,00*4D
1535760001.766 $GPDTM,W84,,00.0000,N,00.0000,E,,W84*41
1535760001.766 $GPGGA,000002,3242.4592,N,11714.2041,W,2,8,2.3,12,M,,M,,*53
1535760001.766 $GPVTG,342.8,T,331.8,M,0.0,N,0.0,K*4A
1535760001.766 $GPZDA,000002,01,09,2018,07,00*4E
...
```

Gotchyas

- We don't always get complete data
- Pad dates of data included by one day to avoid confusion with local time versus UTC
- Changing file naming or subdirectories can lead to us missing the data
- Use appropriate resolution: >1Hz data needs decimal places
- Incomplete strings - checksums can help

R2R
ROLLING DECK TO REPOSITORY

For Ship Operators

Underway Data Profiles

Online Vessel Profiler - requires login

www.rvdata.us/community/ship-operators

Recommended Best Practices

- Cruise Data Directory Structure
- Navigation Data Collection (PDF)
- Sensor Coordinate Systems
- Draft: Underway Transmissometer Best Practices (PDF)

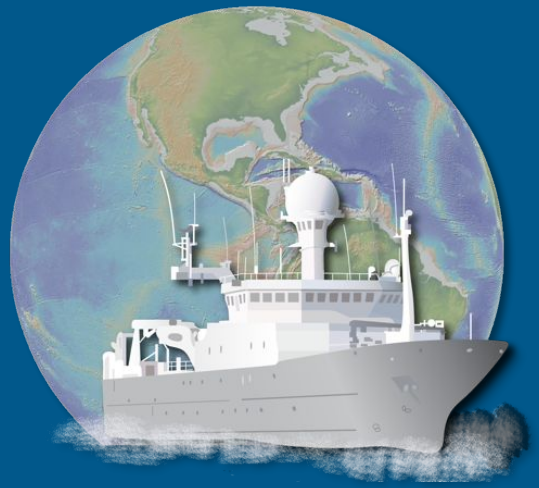
UNOLS/R2R Cruise Personnel Manifest (v4.0 template)

Moving Forward

- Updated best practices
- Converging on common formatting
- We can work with you if you want to make changes

Best Practices for Navigation

- What causes data gaps?
- How can we monitor gps to catch problems earlier?



From Operator to R2R: Improving the Flow Together

New R2R Tools to Help Operators & Final Comments

Suzanne O'Hara
LDEO

R2R - Looking Forward

- Creating new web services and website search functionality
- Extracting & serving instrument documentation from cruise distros
- Working with UNOLS to find ways to share cruise metadata (use UNOLS cruise planner)
- Creating Operator Dashboard and services
- Exploring better ways to exchange device metadata
- Exploring near-realtime data streams
- ??

Want to Join R2R ?

LDEO is looking for a new hire to work within the Geoinformatics Research Group.

Responsibilities include curation of data arriving at R2R from the US Academic Research Fleet and preparation of mulitbeam bathymetry data for inclusion in GMRT.

Ask me for more information

Thank you for participating in this discussion.

Contact us if you have additional questions, comments or suggestions.

Contact us:

- In person during RVTEC
- Email info@rvdata.us
- Visit www.rvdata.us
- @R2RData

Acknowledgements

Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE

WOODS HOLE OCEANOGRAPHIC INSTITUTION 1930

FLORIDA STATE UNIVERSITY 1851

SCRIPPS INSTITUTION OF OCEANOGRAPHY

NSF

Office of Naval Research ONR Science & Technology

NOAA

SCHMIDT OCEAN INSTITUTE

UNOLS