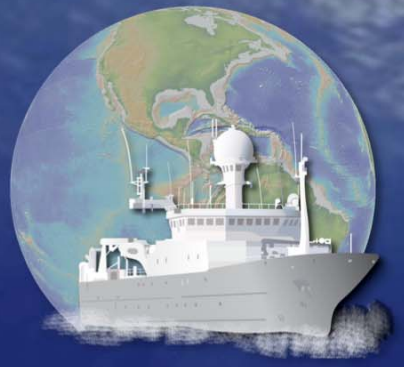

rvdata.us

Rolling Deck to Repository (R2R)

Focus Group:
Cruise Directory Structure

Context

- Proactive guidance for future development e.g. new vessels and/or scheduled upgrades.
- Not required for R2R participation.

http://www.rvdata.us/operators/directory

Design Goals

1. Preserve original “full”-resolution data from the instruments.

2. Organize files in fleet-standard directory structure that can be routinely “broken out” shoreside.

    a) Segregate “routine underway” data from everything else.

    b) Segregate data from documentation.

    c) Segregate the original instrument data from secondary/ processed products.

3. Preserve “exact copy” of what the science party took home.

Input

- In-person/small-group meetings at operating institutions
- Requirements from data managers at NOAA and allied NSF data centers
- Presentations/discussions at national meetings
- Survey of current practice (existing distros) across the fleet
- Convergence with SAMOS/real-time discussion thread

Directory Structure *(current draft)*

cruise-id/

reports/ – personnel list, operations summary, etc.

r2r/

eventlogs/ – produced by, and/or formatted for, R2R Event Logger

devicetype[_[make][model]][_location]/ e.g. *adcp, ctd, multibeam, subbottom, xbt,..*

docs/ – manuals, photos, calibration sheets, etc.

raw/

proc/

(where *devicetype* is a standalone system that logs independently)

system/ e.g. *uhdas, scs, calliope, lds, chud, midas, sms, suds, udas, vids,..*

docs/

raw/

devicetype[_[make][model]][_location][_mnemonic]–dateTtimeZ (file)

proc/

(where *system* is a serial multiplexing/timetagging acquisition system, and *mnemonic* allows further differentiation)

science/ – "dropbox" for files created by science party

photos/ – intended for public release

Best Practices

1. Cruise IDs are unique within R2R. The exact naming style is each operator's prerogative, but IDs should begin with the vessel's unique prefix.

http://www.rvdata.us/voc/vessel

Best Practices *(cont.)*

2. Directory and file names use standard vocabulary of device type, make, and model.

http://www.rvdata.us/voc/devicetype

Names are lower case, with no spaces or non-alphanumeric characters except nonconsecutive underscores and prefix.

R2R will never chop up, recombine, or rename files.

Best Practices *(cont.)*

3. Write all data from the same “talker” (instrument interface) to the same file (i.e. multiple NMEA sentences from same GNSS receiver port).

Data file granularity is daily where possible.

Best Practices *(cont.)*

4. Individual files are never compressed.

   If compression is required, create a tarball of the entire cruise distro and zip that tarball using GNU gzip (.gz.).

   Zip on-the-fly during network transfer.

Best Practices *(cont.)*

5. Operator creates a checksum manifest using the “md5deep” open-source package –

md5deep -c -r -l -o f -t -z cruise > cruise.md5deep

Existence of a manifest indicates distro is complete.

R2R will recalculate and validate the manifest shoreside.

Best Practices *(cont.)*

6. R2R will pull cruise distros from operators via network transfer using “rsync” (requires ssh access).

Alternative for large distros is shipment of USB portable drive. NTFS filesystem is preferred.

Issues

- Do we need a controlled vocabulary for “location” (e.g. *fore*, *aft*, *starboard, port*, *seachest*, *wetlab*, *doghouse*,..)?

- Should we exclude “auxiliary” data sets (created underway for e.g. realtime displays) from online dissemination?

- What are “raw” and “processed”?