Funded by the National Science Foundation, Grant #2127548

Cyberinfrastructure for Major Facilities: Challenge and the Path Forward

by Ken Feldman (UW), Laura Stolp (WHOI),and Rebecca Hudak (WHOI), with input from R2R and ARF

Cyberinfrastructure for Major Facilities: Challenges and a Path Forward

March 2, 2022
CI Compass Cyberinfrastructure for NSF Major Facilities Workshop

USC Viterbi School of Engineering Information Sciences Institute
INDIANA UNIVERSITY
UNIVERSITY OF NOTRE DAME
THE UNIVERSITY OF UTAH
TEXAS TECH UNIVERSITY
renci

Cyberinfrastructure for Major Facilities: Challenges and a path forward

ARF and R2R

Ken Feldman (UW), Laura Stolp (WHOI), Rebecca Hudak (WHOI)
with input from R2R and the ARF

Affiliations

UNOLS

UNIVERSITY-NATIONAL OCEANOGRAPHIC LABORATORY SYSTEM

NSF

Office of Naval Research

ONR

Science & Technology

R2R

Academic Research Fleet

- Vessels are remote research platforms
- Operated by different institutions
- Low bandwidth/ high latency
  - Not all vessels have the same connectivity
  - Some blockage zones depending on heading and/or location
- CI and Data Distribution rules governed by multiple entities
  - Owner Entity
  - Operator Institution
  - PI Institution
  - Grant Funding Institution
- Ship sensor field data are managed through the R2R data facility
  - Common data pipeline and services for all ships

Landscape in which ARF Operates

- University-National Oceanographic Laboratory System (UNOLS)
  - Includes over 60 National Laboratories involved in oceanographic research dedicated to coordinating oceanographic ships' schedules and research facilities.
- NAVY
  - Office of Naval Research (ONR)
- National Science Foundation
  - Geosciences (GEO)-> Ocean Sciences (OCE)
- Individual Operators
  - Home Institution
- Principle Investigator
  - Could be from the International Science community - write the proposal to charter the vessel - expedition
- Grant Funding Agency
  - E.g. NSF, ONR, NOAA, NASA, Other
- R2R
  - Lamont-Doherty Earth Observatory - Columbia
  - Scripps Institute of Oceanography - University of California
  - Florida State University
  - Wood Hole Oceanographic Institution

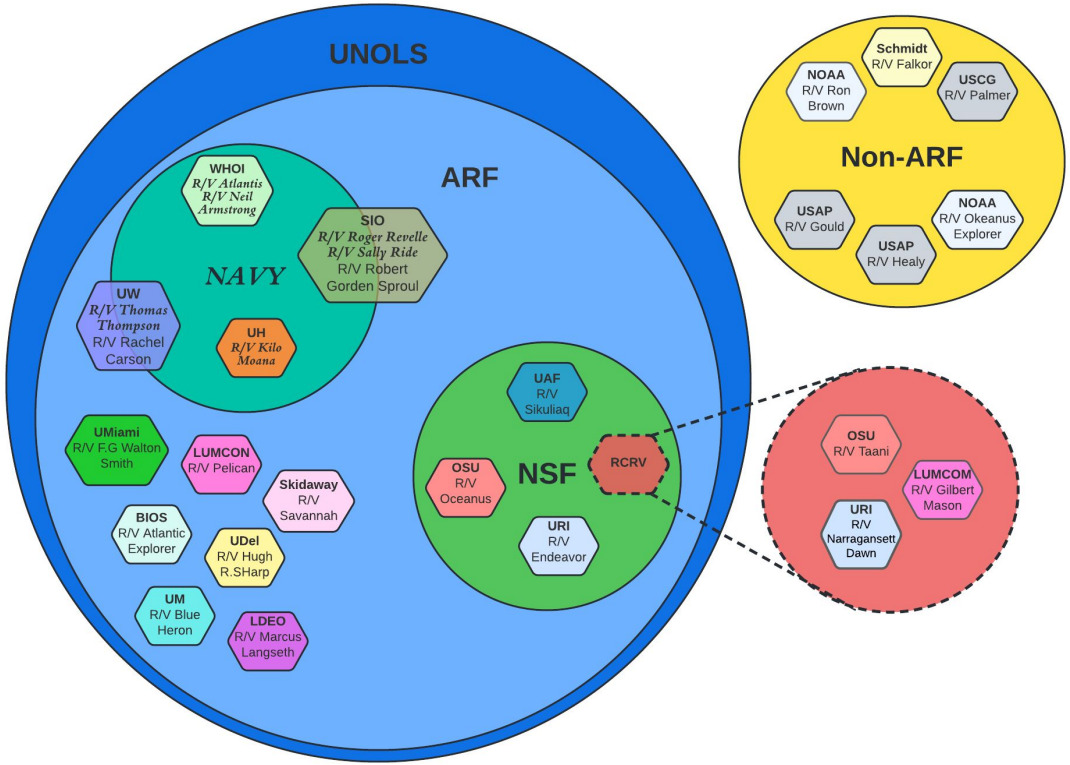
ARF Global and Ocean Class Vessels

Vessel Name	Operator	Owner	SOA	Year Built/midlife	Sci Party
Global Class Vessels					
THOMAS G. THOMPSON	UW	Navy	274/84	1991/2017	36
ROGER REVELLE	SIO	Navy	273/83	1996/2020	37
ATLANTIS	WHOI	Navy	274/84	1997/2021	37
SIKULIAQ	UAF	NSF	261/79	2014	24
MARCUS G. LANGSETH	LDEO	LDEO	235/72	2006	35
Ocean/Intermediate Class Vessels					
KILO MOANA	UH	Navy	186/57	2002	29
OCEANUS	OSU	NSF	177/54	1975/1994	18
ENDEAVOR	URI	NSF	185/56	1976/1993	18
ATLANTIC EXPLORER	BIOS	BIOS	170/52	1982/2006	22
NEIL ARMSTRONG	WHOI	Navy	238/73	2014	24
SALLY RIDE	SIO	Navy	238/73	2014	25

ARF Regional, Coastal and Local Vessels

Regional Class Vessels					
HUGH R. SHARP	Udel	Udel	146/44.6	2005	14
Coastal/Local Vessels					
ROBERT GORDON SPROUL	SIO	UC	125/38	1981/1985	12
PELICAN	LUMCON	LUMCON	116/35	1985/2005	14
F.G. WALTON SMITH	Miami	Umiami	96/29	2000	12
SAVANNAH	Skidaway	SkIO	92/28	2001	16
BLUE HERON	UM-Duluth	Uminn	86/26	1985/1997-98	6
RACHEL CARSON	UW	UW	72/22	2003/2017	9

Graphic Representation of ARF Governance Landscape



Sealink Bandwidth

CIR shore-to-ship	CIR ship-to-shore	MIR shore-to-ship	MIR ship-to-shore	Notes
512	256	1024	512	Nominal plan for vessels with less than 40 berths (Q1 2020 to ~Q1 2022).
1024	512	4096	1024	Nominal plan for vessels with 40 berths or more (Q1 2020 to ~Q1 2022).
2048	1024	6144	2048	–
2048	2048	6144	2048	Nominal plan for 1.5m radome-equipped ARF vessels (beginning Q4 2021-Q1 2022)
4096	2048	8192	4096	Nominal plan for 2.4m radome-equipped ARF vessels (beginning Q4 2021-Q1 2022)

* source: www.hiseasnet.com

Fleet Xpress Bandwidth

CIR shore-to-ship	CIR ship-to-shore	MIR shore-to-ship	MIR ship-to-shore	Notes
512	256	4096	2048	Nominal plan for vessels with less than 40 berths (Q1 2020 to ~Q1 2022).
512	512	8192	4096	—
1024	512	8192	4096	Nominal plan for vessels with 40 berths or more (Q1 2020 to ~Q1 2022).
1024	1024	8192	4096	—
2048	1024	8192	4096	—
2048	2048	8192	4096	Nominal plan for ARF vessels (beginning Q4 2021-Q1 2022)
3072	1536	8192	4096	Subject to capacity approval. **90 days advanced planning recommended**.
3072	1536	8192	4096	Subject to capacity approval. **90 days advanced planning recommended**.
3072	3072	8192	4096	Subject to capacity approval. **90 days advanced planning recommended**.

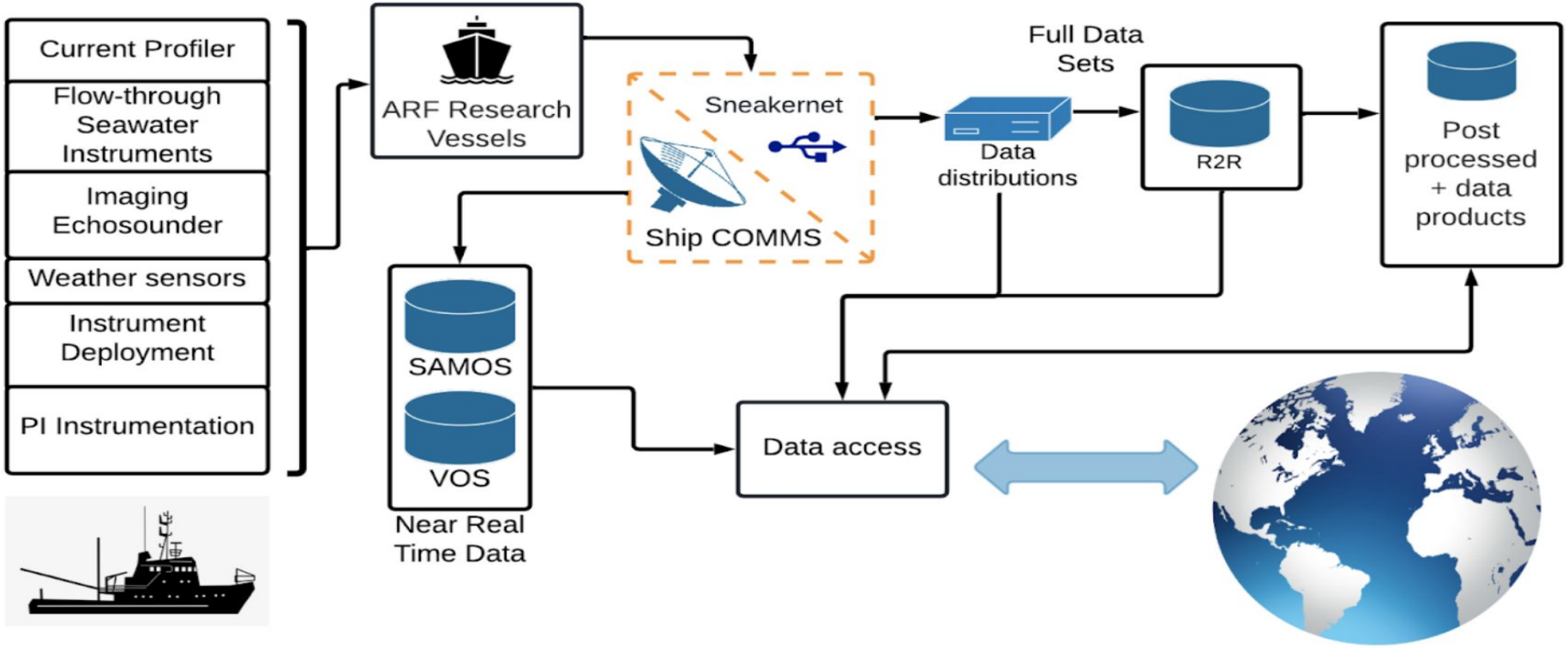
* source: www.hiseasnet.com

Rolling Deck to Repository (R2R)

The Rolling Deck to Repository (R2R) program provides fleet-wide management of underway data to ensure preservation of, and access to, our national oceanographic research assets.

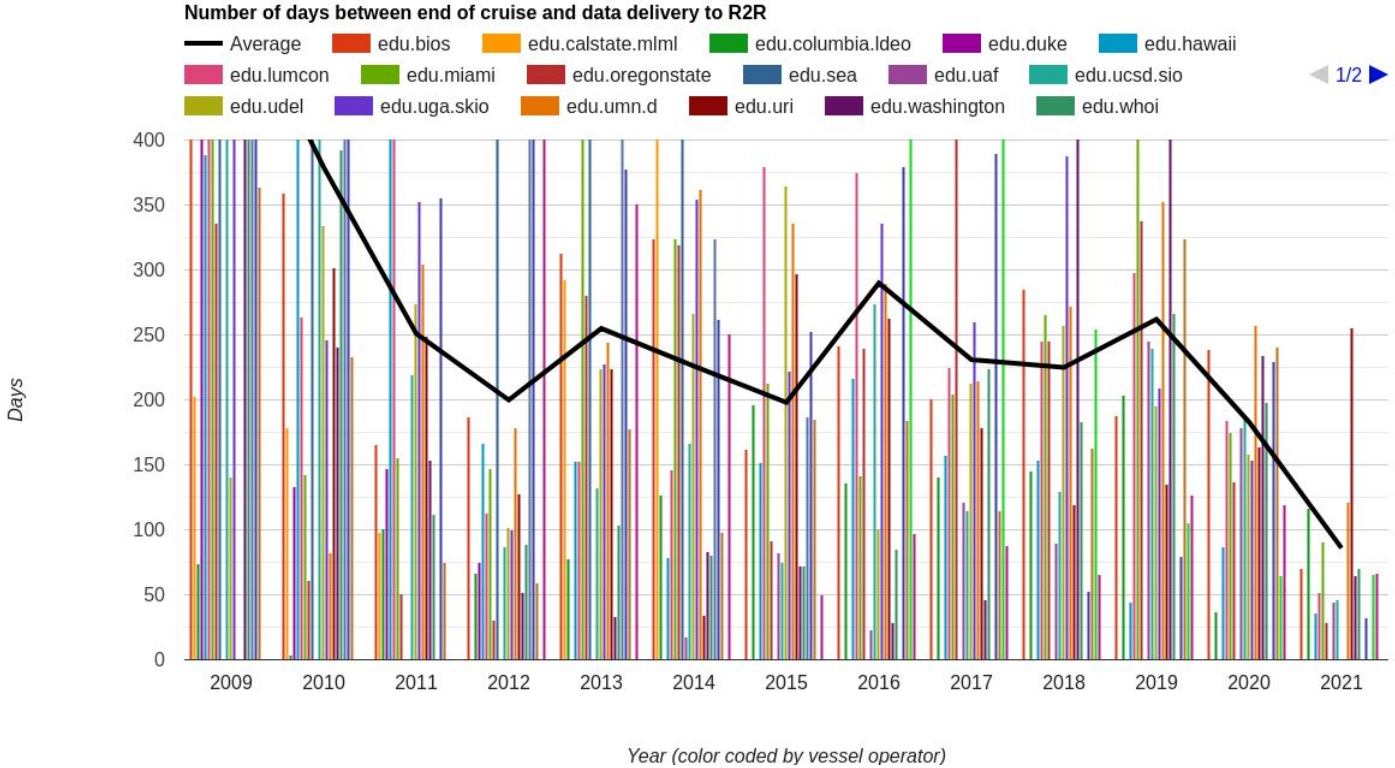
- Post Cruise data distribution are created and submitted to R2R
- Underway meteorological, thermosalinograph, sea temperature data is processed in near real-time on shore through a partnership with SAMOS
  (Shipboard Automated Meteorological and Oceanographic System)
- The raw data distributions are submitted to deep archives in their entirety.
- R2R breaks out cruise distributions and submits to NCEI (National Centers for Environmental Information)
- R2R performs QA and DP on a subset of the data and submits to NCEI
- More information available at https://rvdata.us

Data Flow from Ship



Data Distribution travel time from vessel to R2R

Data valid as of 2022-2-18

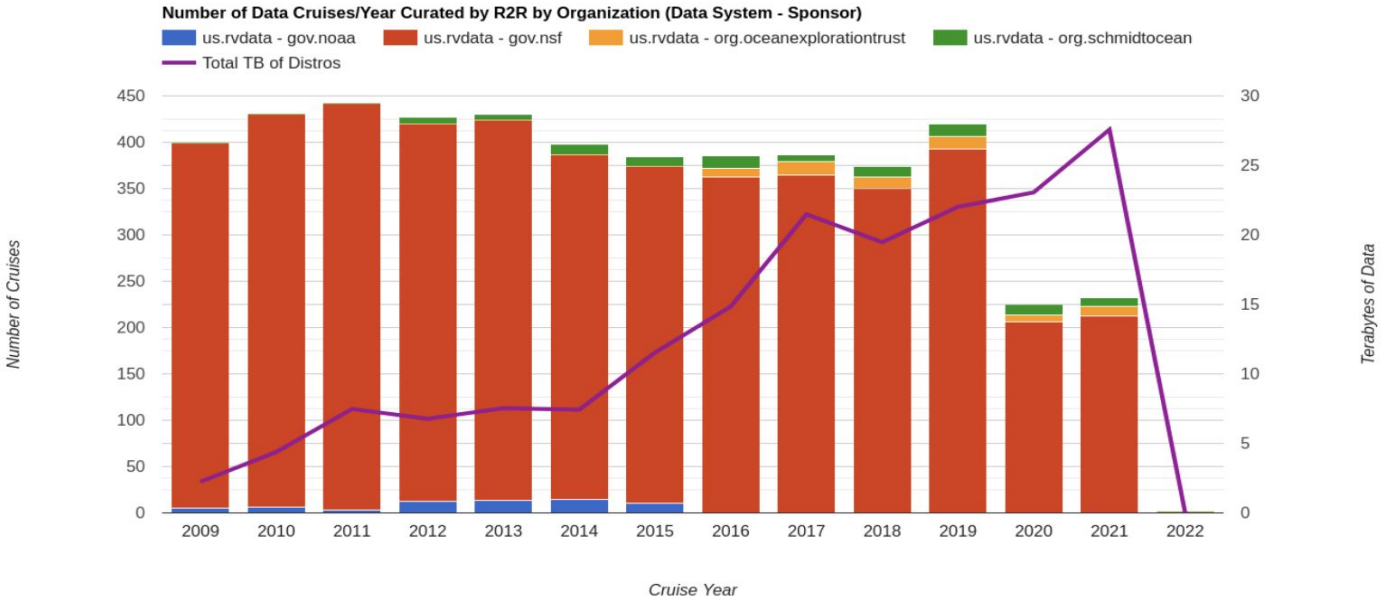


Cruises Curated by R2R

Data valid as of 2022-2-18

Number of Data Cruises for Selected Parameters: 4933

Number of Non-Data Cruises for Selected Parameters: 213



CI Challenges

- CI updates are not yet incorporated as part of normal working flow on many vessels
  - Constraints: Time, Access, Personnel
  - CI Equipment currently competes with instrumentation for funding
- The volume of data and its rate of increase varies by vessel
  - Smaller ships may be about 1GB per cruise
  - Larger may be in the TBs per cruise. Largest was about 6 TB
- The larger datasets are harder to transfer from ship to shore, shore to shore
- Instruments change in that existing are replaced and brand new ones are installed.
- New vessels are being built with newer instrumentation often generating larger datasets, more instruments, and fewer science party personnel
- Ship to Communications are also increasing and becoming more reliable, but not as fast as demand
  - More and more services are reliant on cloud services
- 1-2 staff that have to manage all, often in addition to many other duties

One predominant CI problem that you would like to solve?

- Ability to make collected data available for science party use on shore
  - This will enable science party members to be more productive collaborators from shore without the need for every member to go to sea.
    - I.e. some people can be involved for the 2-3 days their expertise is most effective rather than committing to 30 days at sea
    - Standardized methods for aggregating data and associated metadata prior to transfers to shore

Challenges and Opportunities

- **Potential opportunities to collaborate with other facilities**
  - Utilizing Shared Services offered by the Broader NSF CI Ecosystem
  - Collaborations and/or Sharing Best Practices for CI Between Major Facilities and the Broader CI Ecosystem
  - Balancing adoption of new/emerging technologies with current operations
- **Biggest challenges that overshadow most such considerations for ARF/R2R:**
  - Low bandwidth and high latency satellite communications present unique challenges to using off-ship services
    - e.g. cloud-based services
  - Difficult to devote time and resources to incorporating new technologies and training personnel
    - Most of our personnel operate at sea for long periods of time and juggle multiple responsibilities.
  - Vessels are operated by separate institutions with infrastructure designed and structured independently.
  - ARF in early stages of operating cohesively as a large facility and converging our disparate systems
  - A clear understanding of the benefits of new services along with time set aside for training and development will improve adoption and likelihood of success
  - RCRV's are in the design and implementation phase, will collect more data than previous vessels, and will have fewer science berths than previous global and ocean class vessels.
- **On the plus side**
  - We are engaging ResearchSOC as our virtual CISO. As they come up to speed with the intricacies of ARF they will be able to advise on what services might be available that can work well for us.
  - We are actively interested in learning about any services that can be applicable given our constraints.

Thank You!