Rolling Deck to Repository (R2R) Recommended Data Practices

Presenters: Karen Stocks, George Dubinin, Gwynne Hayes, Rebecca Hudak, Shawn Smith

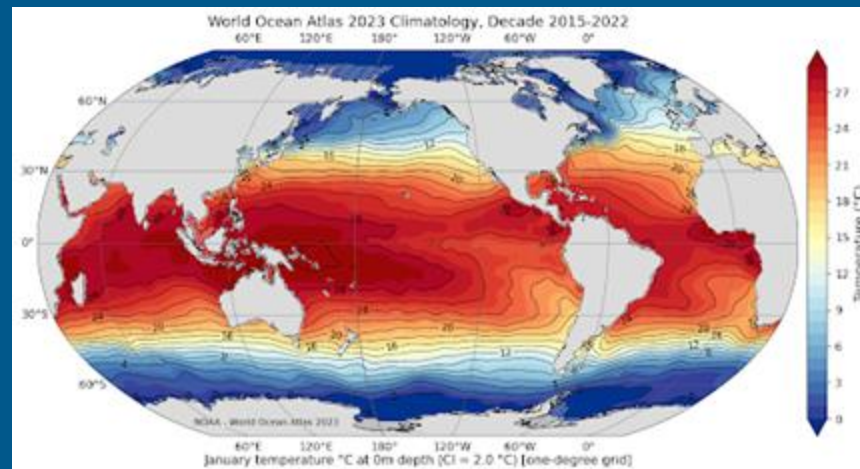# Introducing the R2R team

R2R Overview (?)

Karen Stocks

R2R
ROLLING DECK TO REPOSITORY
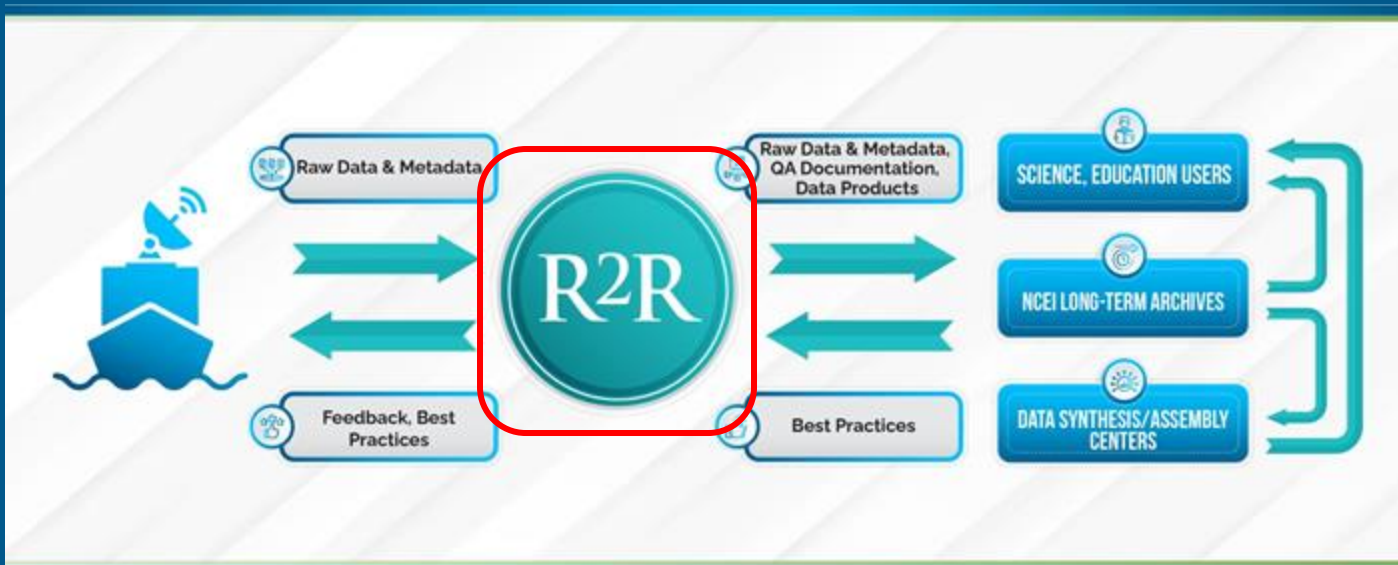
R2R Funding/In-kind Support

Timeline

2008: Pilot

2009 NSF Award

2014 NSF Award

2019 NSF Award

2025 NSF Award (5 yr)

Mission

To preserve and provide access to underway data from the US Academic Research Fleet

R2R
ROLLING DECK TO REPOSITORY

Global Products

Earth Magnetic Anomaly Grid (EMAG2v3)
2 arc-minute resolution
4 km Altitude

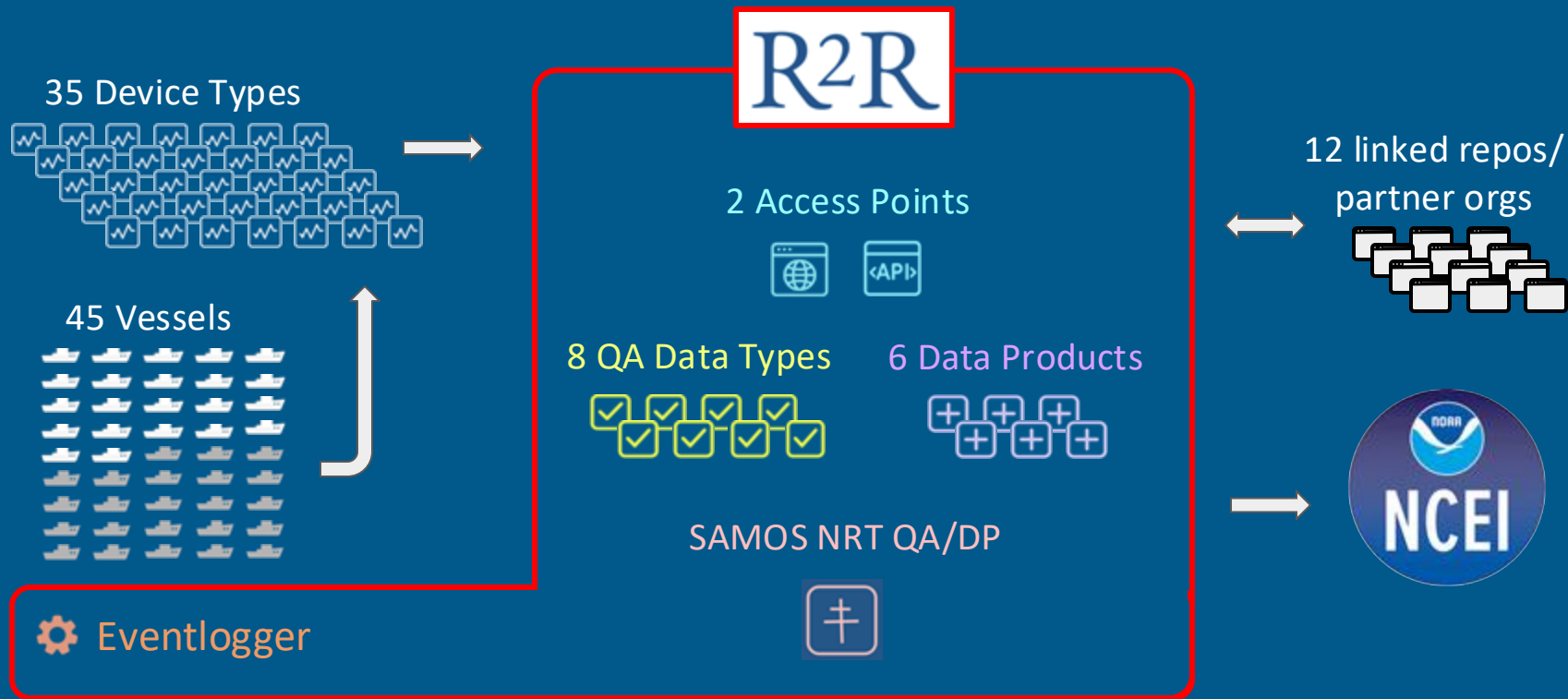World Ocean Atlas 2023 Climatology, Decade 2015-2022

Create Select Quality Assessment/Data Products
Support Discovery and Access
Manage Data Releases
Develop and Maintain Cyberinfrastructure

R2R By the Numbers

35 Device Types

45 Vessels

R2R

2 Access Points

8 QA Data Types   **6 Data Products**

SAMOS NRT QA/DP

⚙ Eventlogger

12 linked repos/ partner orgs

NCEI

5,800+ cruises; 60,000+ filesets, 23M+ downloadable files, ~350 TB single copy

R2R
ROLLING DECK TO REPOSITORY

SAMOS: Near Real Time (NRT) Met & TSG
Eventlogger: NRT science party & technician events metadata

SAMOS - Near Real Time MET and TSG Evaluation

History

- Providing high-quality underway meteorological and oceanographic data from research vessels (RV) to the scientific community since 2005

Users

- Satellite algorithm developers
- Air-sea exchange (flux) researchers
- Ocean and atmosphere modelers
- Operational forecasters
- Geoinformatics, Ocean Best Practices

Status

- 30 vessels active in 2025
- 9.3 million 1-minute data records processed in 2024
- Global coverage, concentrated around North America.

Agenda

- Recommended Data Practices

- Event Log Harvesting

- R2R and FAIR Principles

Recommended Data Practices

George Dubinin, Gwynne Hayes

R2R
ROLLING DECK TO REPOSITORY

Regularly Maintain Device Format Documentation

- Device format descriptions describe columns, units, standards, and other features of data files.

- They are important for both R2R and end users.

- For a detailed description and examples see the Data Submission page on the R2R website, www.rvdata.us.

R2R
ROLLING DECK TO REPOSITORY

Metadata Sharing

- R2R needs cruise metadata to begin processing a cruise including:

  - cruise ID, chief scientist, departure/arrival ports, dates, and funding source information.

- The preferred way to submit metadata is entering it into the MFP (Marine Facility Planning) system.

- If you do not use the MFP please reach out to us for guidance.

R2R
ROLLING DECK TO REPOSITORY

Separating Oversize Filesets

- An oversize fileset is any distro subdirectory that is over 500gb.

- These directories should be separated and staged independently from the base distro.

R2R
ROLLING DECK TO REPOSITORY

File-Level Recommended Practices

- Ensure your distro contains a deep checksum file. These files:

  - list all file checksums and paths of the distro.

  - can be generated by acq system or the md5deep linux tool.

- Include (YYYY-MM-DD) date strings in filenames.

- Remove non-ascii characters and quotes from filenames.

R2R
ROLLING DECK TO REPOSITORY

Data Submission and Update Requests

- R2R should receive data distros on at least a quarterly basis.
- Please send complete and finalized distros.
    - It is difficult to make changes to data once they have been processed at R2R.
    - If changes are needed, please reach out to us at info@rvdata.us before sending a new distro.

R2R
ROLLING DECK TO REPOSITORY

Network Transfers

- Our preferred method of network transfer is Globus.
  - It performs checks on all transmitted data to ensure file integrity remains uncompromised.
  - It ensures security of the transfer.
  - Pushing data to R2R doesn't require a subscription.
- If you are pushing a large distro, please let us know before via info@rvdata.us.
- We are happy to help you set Globus up!

R2R
ROLLING DECK TO REPOSITORY

Water Column Sonar Data

- Confirm water column sonar data contains navigation before sending it to R2R.

R2R
ROLLING DECK TO REPOSITORY
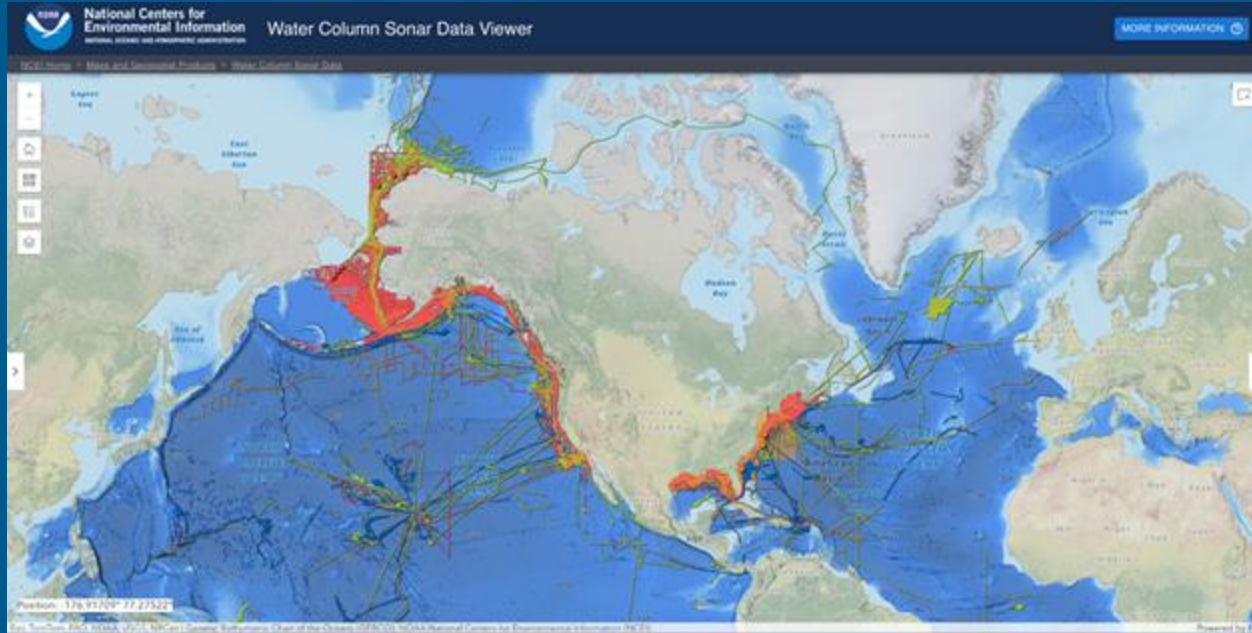
Gravity Ties and Magnetometer Laybacks

- Include gravity ties and magnetometer laybacks in every distro.
  - Store these files in a consistent directory with a consistent naming structure.
    - EX: gravtie_<date>.pdf
- Do not send us anything that is unable to be released.
- PFPE's Github page has more information regarding reading and processing gravity data.

For more general gravity information refer to PFPE's Github

R2R Data Submission

For more information refer to the Data Submission page of rvdata.us
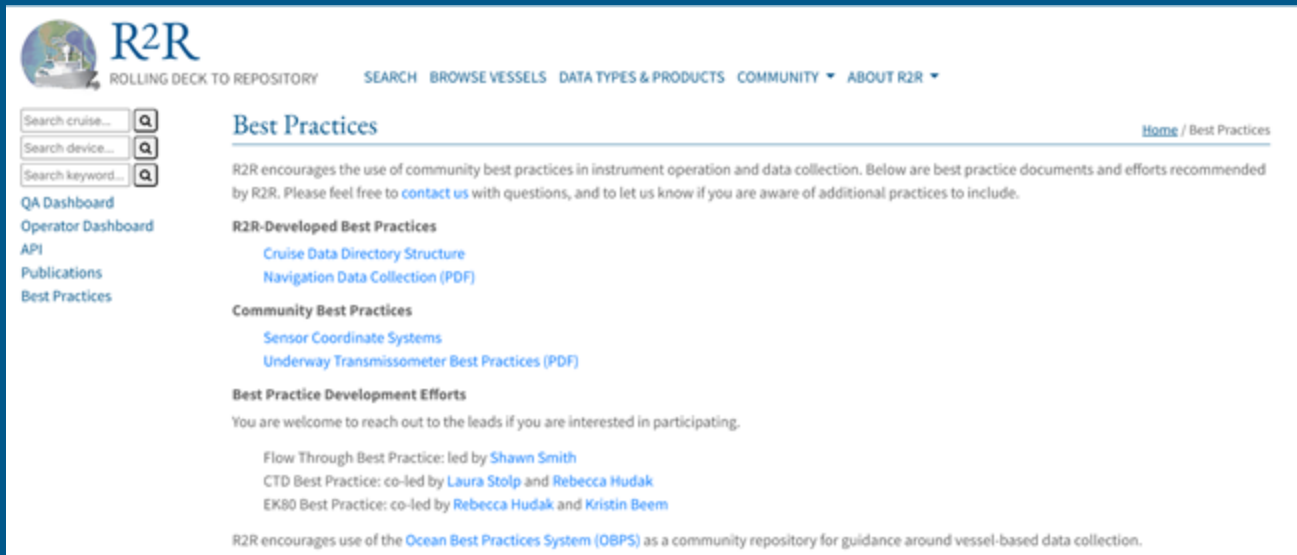
https://www.rvdata.us/about/data-policies-and-repositories/data-submission

R2R Recommended Practices

For more information refer to the Best Practices page of www.rvdata.us

https://www.rvdata.us/recommended_practices

EventLogger Harvesting

Becca Hudak

At-Sea Event Logging - Science and Technician Events

"EN709", "R/V Endeavor", "", "", "", "", "", "", "", Page 1 of 10

List | Find | Help

Summary | Threaded   -- Author -- v   -- Instrument --   v -- Action --   192 Entries

Goto page 1, 2, 3 ... 8, 9, 10   Next All

Event	dateTimeUTC	GPS_Time	Instrument	Action	Transect	Station	Cast	Latitude	Longitude	Seafloor	Author	Comment
20230904.1303.001	20230904.1303	2023/09/04 13:03:49	Ship	startCruise	NaN	NaN	NaN	41.523698	-70.672452		lButler1	Lines off from WHOI dock
20230904.1303.002	20230904.1255	2023/09/04 12:55:14	Ship	other	NaN	NaN	NaN	41.523795	-70.672260	16 m	lButler1	Gangway removed from dock at WHOI and 01 deck to storage on board
20230904.1305.001	20230904.1305	2023/09/04 13:05:36	MeteorologicalSensor	other	NaN	NaN	NaN	41.523220	-70.672742	16 m	lButler1	met sensors operating and logging as usual
20230904.1305.002	20230904.1305	2023/09/04 13:05:59	Echosounder12	start	NaN	NaN	NaN	41.523095	-70.672790	16 m	lButler1	12 KHz started just after leaving the dock
20230904.1324.001	20230904.1454	2023/09/04 14:54:04	UHDAS for both ADCPs	start	NaN	NaN	NaN	41.352133	-70.922173	26 m	lButler1	UHDAS controls both OS75 and WH300 KHz ADCPs. Dan Torres modified the settings. Started south of Cuttyhunk.
20230904.1459.001	20230904.1230	2023/09/04 12:30:04	Other	other	NaN	NaN	NaN	41.523793	-70.672267		lButler1	Pre-cruise 10 minute orientaion
20230904.1500.001	20230904.1400	2023/09/04 14:00:44	Other	start	NaN	NaN	NaN	41.437463	-70.782648		lButler1	1.5 hour safety orientation video and talk by 2nd Mate
20230904.1951.001	20230904.1941	2023/09/04 19:41:49	Fluorometers on underway impeller	clean	NaN	NaN	NaN	41.779095	-70.487005		bClarke1	clean fluorometers before starting impeller pump flow through system and turning them on
20230904.2014.001	20230904.2002	2023/09/04 20:02:04	Other	start	NaN	NaN	NaN	41.786663	-70.442190		anOther	bag transfer
20230904.2014.002	20230904.2015	2023/09/04 20:15:19	Thermosalinographs on underway impeller	start	NaN	NaN	NaN	41.802885	-70.433955	24 m	lButler1	started sci sw impeller pump. SBE45 and SBE21 with fluorometers.
20230905.2233.001	20230905.1430	2023/09/05 14:30:39	Other	other	NaN	NaN	NaN	42.671317	-69.059072		lButler1	CTD training
20230907.1247.001	20230907.1248	2023/09/07 12:48:04	Thermosalinographs on underway impeller	other	NaN	NaN	NaN	43.339933	-61.930870		bClarke1	Turn off flow to SBE21 for maintenance. SBE21 salinity reading ~0.1PSU lower than value from SBE45. Stop SeaSave acq for SBE21
20230907.1251.001	20230907.1200	2023/09/07 12:00:49	Thermosalinographs on underway impeller	other	NaN	NaN	NaN	43.338192	-62.115668		lButler1	increase and decrease flow to possibly correct an offset in conductivity values between the SBE21 and SBE45
20230907.1351.001	20230907.1324	2023/09/07 13:23:59	Thermosalinographs on underway impeller	other	NaN	NaN	NaN	43.341535	-61.795250		bClarke1	Start Seasave acquisition file 07Sept2023a.hex after Triton-X and freshwater rinse of SBE21
20230907.1354.001	20230907.1328	2023/09/07 13:27:59	Thermosalinographs on underway impeller	other	NaN	NaN	NaN	43.341657	-61.780355		bClarke1	SBE21 salinity reading ~.3PSU lower than SBE45. Stop SBE21 Seasave acquisition file for cleaning.
20230907.1356.001	20230907.1344	2023/09/07 13:43:59	Thermosalinographs on underway impeller	other	NaN	NaN	NaN	43.342480	-61.720945		bClarke1	Shell was removed from SBE21 during cleaning. Start file 07Sept2023b.hex. SBE21 and SBE45 values agree again.
20230907.1436.001	20230907.1437	2023/09/07	UHDAS for both ADCPs	other	NaN	NaN	NaN	43.342928	-61.518758	82 m	lButler1	Stop acq temporarily to turn Bottom Tracking off for wh300 and os75.

Harvesting of the R2R Eventlogger

- Started working on this initiative this year with SAMOS

- Harvesting Eventlogger data may allow SAMOS to

    (a) maintain up-to-date metadata (e.g., sensor swaps, cal changes) and

    (b) improve data flagging/QC by noting elog documented sensor failures or sensor   impacting events (e.g., icing, roosting birds, etc).

- After several discussions and meeting with Coriolix group we wanted to poll the RVTEC community about an addition of a **SENSOR LOG** to the R2R Eventlogger

R2R
ROLLING DECK TO REPOSITORY

Harvesting of the R2R SENSOR LOG?

- Utilize the R2R Eventlogger controlled vocabulary for instruments BUT;
    - With actions specific for sensors (malfunction, replaced, repaired, cleaned, calibrated)
    - With serial numbers related to sensors
    - A continuous log that does not stop at the end of a cruise
    - In talks with NERC-BODC group about location controlled vocabulary

- Accessible via the R2R Eventlogger, a separate tab on the main page [localhost]:8090

Sensor Log Poll

When you have a minute please take this poll to help give us a better idea on how you use sensor logs!

R2R
ROLLING DECK TO REPOSITORY

R2R FAIR Data Working Group

Shawn R. Smith

R2R
ROLLING DECK TO REPOSITORY

R2R FAIR Data Working Group

- R2R is convening a working group to develop guidance for the ARF to align underway data collection, documentation, and distribution practices with the FAIR Principles

Summary of Principles

- The Findability (F) principles focus on making data and digital resources discoverable.
  - To achieve this, data should be assigned globally unique and persistent identifiers (such as DOIs or URIs), cataloged, and described using rich metadata for search and discovery.

- The Accessibility (A) principles emphasize that data and digital resources should be accessible with persistent metadata through an open, free, standardized communication protocol that allows for authentication and authorization procedure when appropriate.

Summary of Principles

- The Interoperability (I) principles aim to facilitate seamless integration and exchange of data and digital resources across different systems, tools, services, and research domains.
  - To achieve this, it suggests that data and digital resources should be structured using standardized and widely accepted data models, formats, and vocabularies, including references to other relevant data and metadata, for example, a reference to utilized metadata standards.

- The Reusability or R principles promote the creation of data and digital resources that can be readily understood, interpreted, and (re)used.
  - This involves adhering to domain-relevant standards and providing comprehensive data documentation, including clear data usage license and detailed provenance.
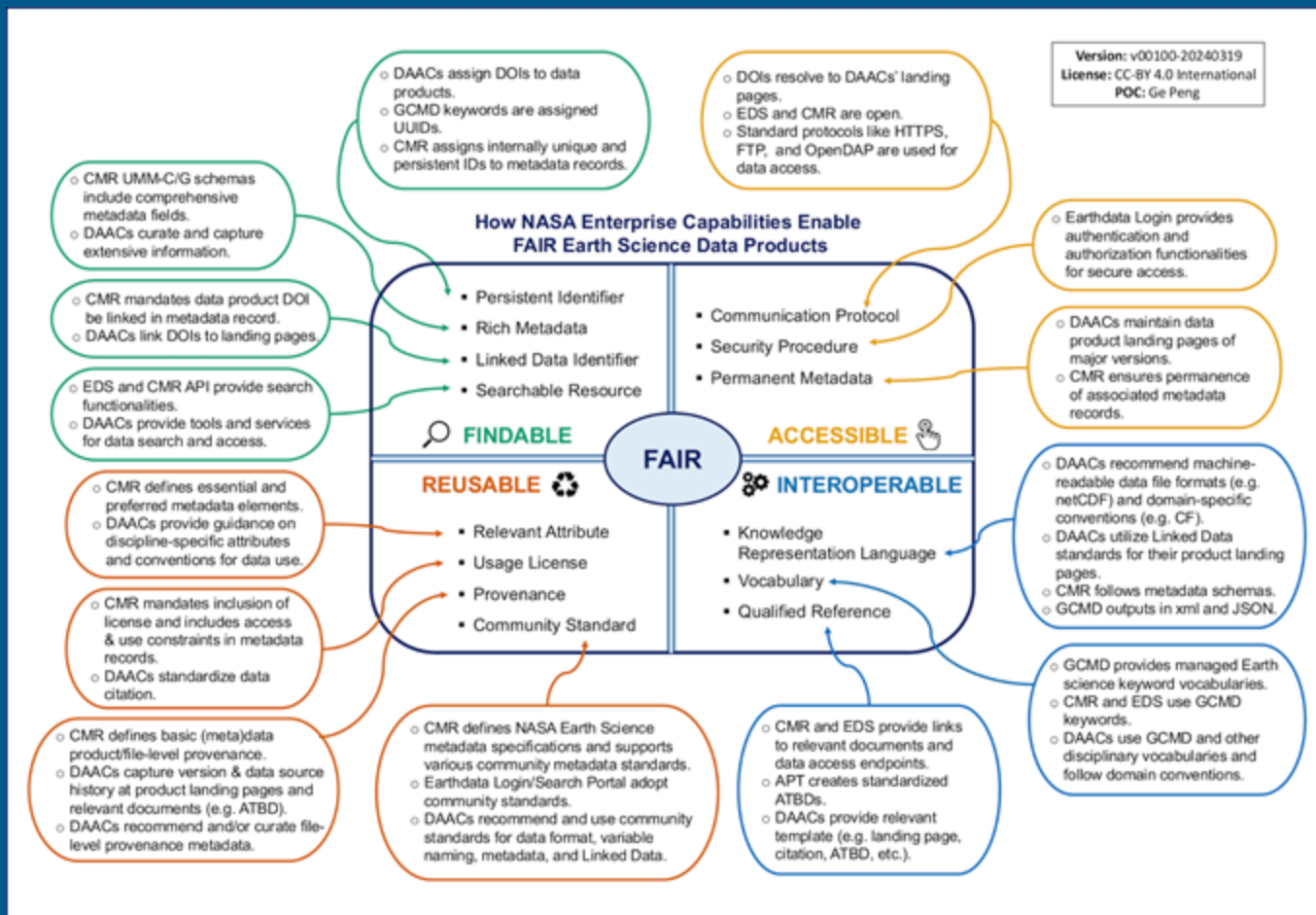
Reproduced from Peng, G., R. R. Downs, H. K. Ramapriyan, Y. Wei, B. Ramachandran, M. Parsons, Z. Liu, and NASA O'FAIR WG, 2024: A Practical Guide for Open, Free & FAIR NASA Earth Science Data Products. Document ID: NASA-OFAIR-ESDSWG-DOC-0002. Version: v00r05-20240614. CC0 1.0 + Attribution. https://doi.org/10.5067/DOC/ESCO/ESDSWG-0002V1

So What Does This Mean for the ARF?

- FAIR Principles are guidelines
  - But how should they be applied to ARF underway data?
  - At what stage in the data lifecycle does FAIR apply?
  - What should be done during data collection to support FAIR principles?
  - What should be handled by R2R and data archivists?
  - What FAIR practices can be implemented at data collection to assist data users, R2R, etc.

- Getting from the FAIR principles to actionable practices is a process
  - Several have gone before us, so examples exist
  - Need a community effort over several years to get us there

R2R
ROLLING DECK TO REPOSITORY

Example from NASA O-FAIR Working Group

How NASA Enterprise Capabilities Enable FAIR Earth Science Data Products

Image source: Peng et al. (2025; DSJ, *under review*).
Preprint: https://doi.org/10.5281/zenodo.15706527

Example: O-FAIR checklist

Check 7. Data is stored in a standard, open, and machine-readable format, following community metadata conventions.

Description: NASA ESCO maintains a list of approved machine-readable data formats, such as netCDF.

Notes: Use CF/ACDD metadata conventions

Responsible Party: DP

Principle(s) Supported: I1-REQ-D, I2-REQ-M, R1.3-REQ-D

Compliance Requirements:

(i) Req 1. Information on data format type is provided or detectable and compliant with GCMD data format keyword.

(ii) Req 2. Data format type is ESCO-approved, machine-readable.

Implementation Notes:

- Machine-readable: structured and parsable (validating syntax follows formal rules, such as JSON, XML, CSV, or NetCDF)

- To validate if it is machine-readable, it should check whether the file is parsable, which might not be feasible/efficient with a large data file.

R2R
ROLLING DECK TO REPOSITORY

R2R FAIR Working Group

- Charge to WG
  - Review existing FAIR implementation plans and procedures in the geosciences
  - Assess the FAIR readiness for ARF underway data using existing assessment tools
  - Develop device-specific plans moving the ARF towards FAIR at the time of data collection
  - Develop guidelines and protocols to align ARF underway data with documented principles.

R2R
ROLLING DECK TO REPOSITORY

R2R FAIR Working Group

- Recruitment for WG is now underway
- Composition of WG
  - R2R team members
  - Marine technicians – Looking for 3-4 who are interested
  - Data managers and archivists
  - FAIR experts
- Interested in contributing to this effort?
  - Please contact Shawn Smith (srsmith@fsu.edu) or reach out to anyone from the R2R team during RVTEC.
- Honorariums are available for some contributors to the WG, including marine technicians.

R2R
ROLLING DECK TO REPOSITORY

Questions for the Community:

- What is your biggest challenge when working with R2R?

- Is there any extra guidance that R2R can provide based on our requests? Other ways we could help?

- Any other feedback?

Thank you

info@rvdata.us

Providing access to and ensuring
the preservation of national
oceanographic research data.

R2R
ROLLING DECK TO REPOSITORY