4th International Marine Energy Data Sharing Workshop Report

Wednesday, May 8, 2024
8:00 – 10:00 AM PDT (3:00 – 5:00 PM UTC)
Jon Weers, Andrea Copping, Cesar Castillo, Hayley Farr, Dominick DeCanio

Summary

On behalf of Ocean Energy Systems (OES) and the U.S. Department of Energy, the Portal and Repository for Information on Marine Renewable Energy (PRIMRE) team, led by the National Renewable Energy Laboratory (NREL), Pacific Northwest National Laboratory (PNNL), and Sandia National Laboratories (Sandia), hosted the fourth international workshop to explore marine energy data sharing on May 8, 2024. Following successful online workshops in 2021, 2022, and 2023, the 2024 workshop featured updates from the PRIMRE team, a presentation on The Crown Estate's Marine Data Exchange, and group discussions focused on data sharing principles, potential applications of machine learning, and remaining barriers and opportunities to marine energy data sharing (Appendix A). A total of 31 marine energy professionals from 8 countries participated in the workshop, including Canada, Ireland, Mexico, Portugal, Singapore, Spain, the United Kingdom, and the United States. The 3 breakout groups were not recorded, but a record of the written and oral discussion was made using notes and Google Jamboards to contribute to this report (Appendix B).

Workshop Structure

After Andrea Copping (PNNL) introduced the purpose of the workshop, Jon Weers (NREL) presented an update on PRIMRE, which consists of seven Knowledge Hubs and several other tools and resources intended to support the global marine energy community. Jon highlighted recent updates to the Marine Energy Atlas, Telesto, and Marine Energy Software, specifically.

Jon also discussed the evolution of sharing science, the existing FAIR (Findable, Accessible, Interoperable, Reusable) data principles, and the recently introduced FARR data principles (FAIR in machine learning, Artificial Intelligence [AI] Readiness, and Reproducibility).

Chelsea Bradbury (The Crown Estate) presented on the Marine Data Exchange, which hosts marine industry survey data, research, and evidence in the United Kingdom. Chelsea also

highlighted the Marine Data Exchange's use of the Marine Environmental Data and Information Network (MEDIN) guidelines and metadata standards, and machine learning applications.

Before splitting into breakout groups, Jon also highlighted how PRIMRE is making marine energy data programmatically accessible through data lakes on the Marine Hydrokinetic Data Repository (MHKDR), the Marine Energy Data Pipeline, and PRIMRE Centralized Search Application Programming Interfaces (APIs).

Three breakout rooms brought workshop participants together, with PRIMRE staff acting as facilitators and notetakers, to discuss data sharing principles, potential applications of machine learning, and remaining barriers and opportunities to marine energy data sharing. Each group was asked to work through five group discussion questions:

1. Are there ongoing barriers and/or opportunities for developing marine energy data systems for your country?
2. Does your country mandate or encourage FAIR standards?
3. What do you know about FARR? Is it something you will include in your systems?
4. Are you considering machine readability for your data?
5. What issues are you having accessing data, archiving data, finding places to put data, on marine energy?

After the breakout sessions, the group facilitators reported their findings to the larger group and discussed next steps with workshop participants.

Key Workshop Discussions

The results of the breakout group discussions are summarized below:

Are there ongoing barriers and/or opportunities for developing marine energy data systems for your country?

- Open Data Incentive Problems
    - Lack of incentives for data generating groups to share results
    - Pressure to monetize data increases sharing barrier
    - Cultural resistance to data sharing
    - Lack of enforcement on data sharing mandates
- Resource Scarcity
    - Funding is a primary hurdle for developers
    - It is difficult to ensure data quality despite the harshness of the marine environment causing instrument failures, and high instrument costs
- Standards
    - Lack of standards impedes data reuse
    - Widespread use of standards would increase study reproducibility

Does your country mandate or encourage FAIR standards?

- Established Usage
    - Standards have been beneficial for continued development
    - FAIR helps define the “how” for dissemination of publicly funded data
    - Many US government agencies require FAIR
    - Data guidelines / preferred metadata schemas ensure quality and reusability
    - Funding and awareness can limit reach, despite mandates supporting FAIR
- Moving Towards FAIR
    - Many countries encourage open data access, though widespread adoption is a work in progress
    - In many cases countries have similar goals but their initiatives and guidance cover a subset of FAIR standards, and possibly requirements beyond FAIR
    - Currently accessible open data is insufficient for many user needs
    - Data might be requested based on other criteria because FAIR data can be slow
- On The Horizon
    - Many countries are beginning to move towards FAIR standards

What do you know about FARR? Is it something you will include in your systems?

- Participant Backgrounds
    - Some participants were well versed in FARR, whereas others first learned of the topic in this session
    - Some participants expressed their willingness to act on AI readiness, whereas others stated their groups were not there yet
    - Excitement at the prospect of introducing AI into this space
- Standards and Reproducibility
    - Consensus that standards are critical
    - Interest in scientific presentations including code, such as a scientific jupyter notebook, to increase reproducibility
    - Interest that FARR may promote adoption of standards

Are you considering machine readability for your data?

- Quality Control
    - Desire for controlled vocabularies to aid cross-institutional knowledge building
    - Structured and hierarchical data types are helpful to aid metadata delivery
    - Data quality will impact machine learning quality - QC is critical
    - Important to ensure quality does not come at the expense of open source adherence
- Responsibility
    - Introduce machine readability within data generation
    - Include data management section within proposal stage of research
    - Promote documentation of data in peer reviewed publications

What issues are you having accessing data, archiving data, finding places to put data, on marine energy?

- Data Sharing and Trust
    - Companies rely on their data and are therefore averse to data sharing

    - Few government datasets available, and these are not well known
    - Stakeholders question trustworthiness of corporate data storage providers
- Discoverability and Access
    - Many participants find it difficult to know what data and tools are available
    - Data of interest may not exist
    - Data moratoriums can prevent access

Conclusion & Next Steps

The fourth international workshop to explore marine energy data sharing demonstrated that while interest remains strong, there continue to be many challenges to data sharing, ensuring data quality, and adequately funding these efforts. The wave of machine learning and AI applications entering into the marine energy space and other sectors will create new challenges to data management, but will also open up vast new opportunities.

The workshop attracted several data practitioners who have participated in previous workshops, as well as a large number of new participants, underscoring an increasing interest in marine energy and data sharing around the world. It also appears that the topic of data sharing and use of standardized formats and principles continues to expand. To that end, the PRIMRE team plans to host a fifth international workshop in 2025 to continue this collaborative effort.

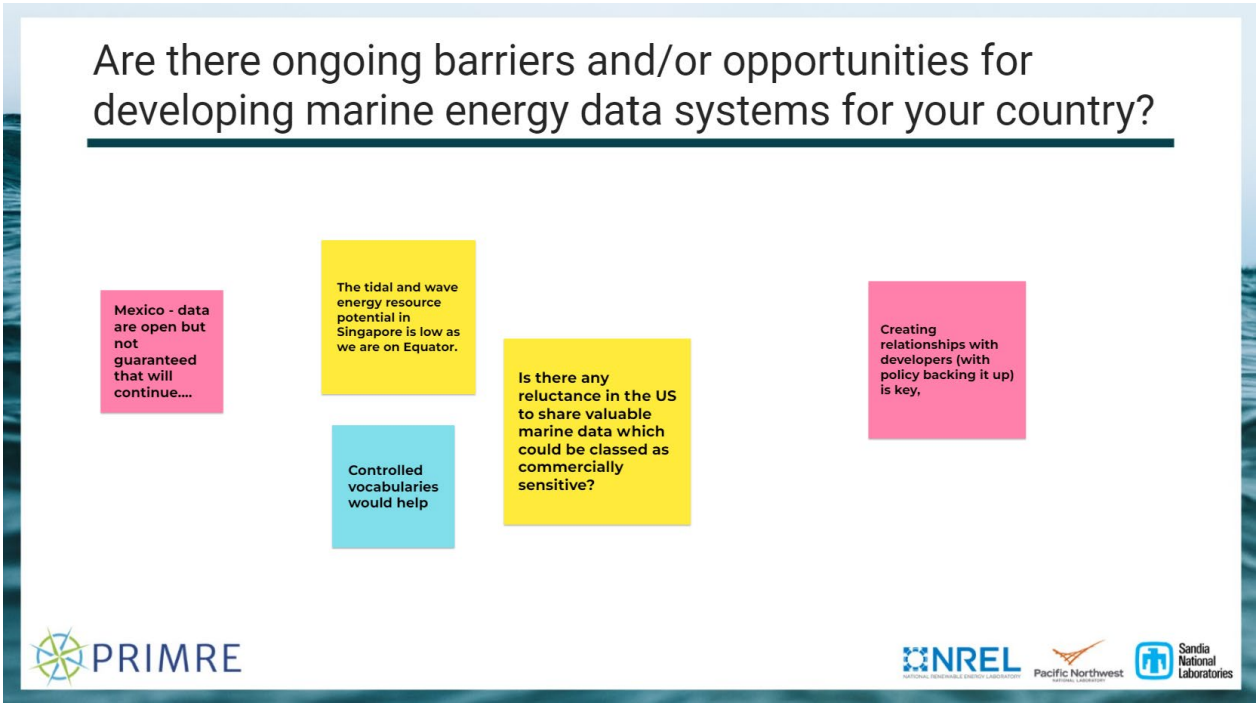
The workshop slides and a recording of the presentations on PRIMRE, Marine Data Exchange, data sharing principles (e.g., FAIR, FARR), and machine learning are publicly available online. A shared Google Drive folder is also available with all workshop materials, including a “living” list of marine energy databases around the world, their scope and format, and their points of contact for all to use and keep up to date.

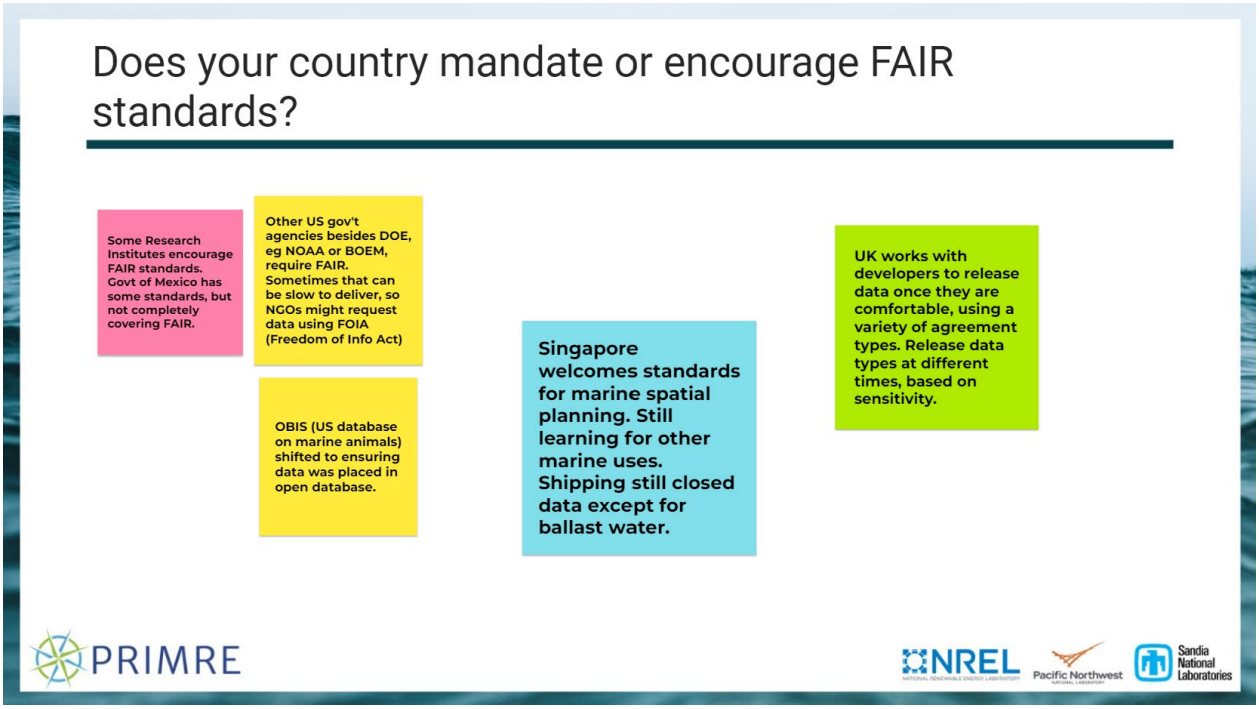
Appendix A – Workshop Agenda

Time	Activity
3:00 - 3:10 PM UTC	Introduction/Purpose of the Workshop
3:10 - 3:25 PM UTC	PRIMRE Updates & Introduction to Data Principles
3:25 - 3:45 PM UTC	Marine Data Exchange Presentation
3:45 - 4:00 PM UTC	Data Sharing & Machine Learning Presentation
4:00 - 4:45 PM UTC	Breakout Group Discussions ● Breakout Group 1 ○ Facilitator: Andrea Copping (PNNL) ○ Notetaker: Curtis Anderson (PNNL) ● Breakout Group 2 ○ Facilitator: Jon Weers (NREL) ○ Notetaker: Austin Venhuizen (NREL) ● Breakout Group 3 ○ Facilitators: Cesar Castillo (Sandia), Jonathan Whiting (PNNL) ○ Notetaker: Megan Anderson (Sandia)
4:45 - 5:00 PM UTC	Report Out & Next Steps

Appendix B – Jamboards

Breakout Group 1 (Andrea)





What do you know about FARR? Is it something you will include in your systems?

Good direction to go, but most countries are not there yet.....most important that we have open science.

Want to see more scientific notebooks like Jupyter and Quarto with code and reproducibility

UK looking to use AI and make data machine useable.

TC114 meeting talked about creating standards as a possible future use of Chatgpt enabled.

PRIMRE

NREL Pacific Northwest Sandia National Laboratories

Are you considering machine readability to your data?

Controlled vocabularies. PRIMRE crosswalk with https://medin.org.uk/data-standards/controlled-vocabularies ?

In Request For Proposal stage, require Data Management section with reference to Best Practices like https://dataoneorg.github.io/Education/bestpractices/

Difficult to make this happen. Who should bear the responsibility....best closest to the source of the data. Dilution once away from the source.

Good idea to promote "data in brief" in peer review publications, helps to guarantee quality.

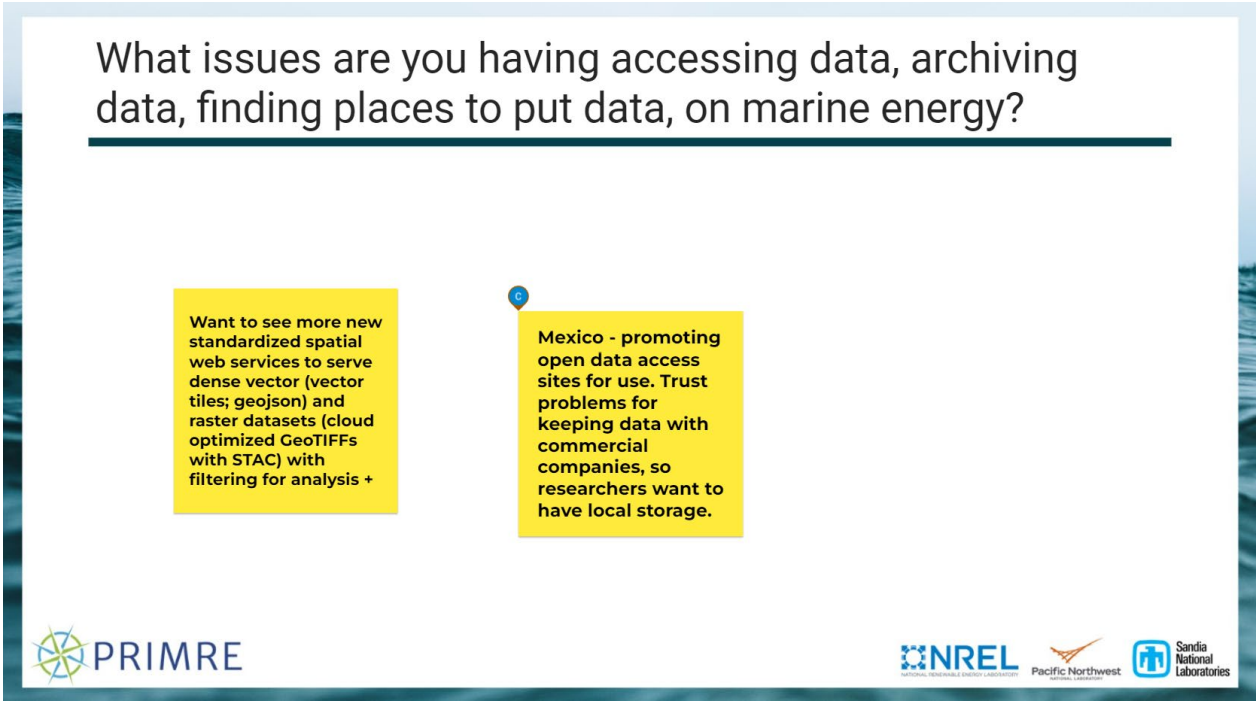
Importance that data are QC'd to ensure that the machine learning is not contaminated with bad data.

Use of specific software to ensure quality....this can be the enemy of open source, esp software.

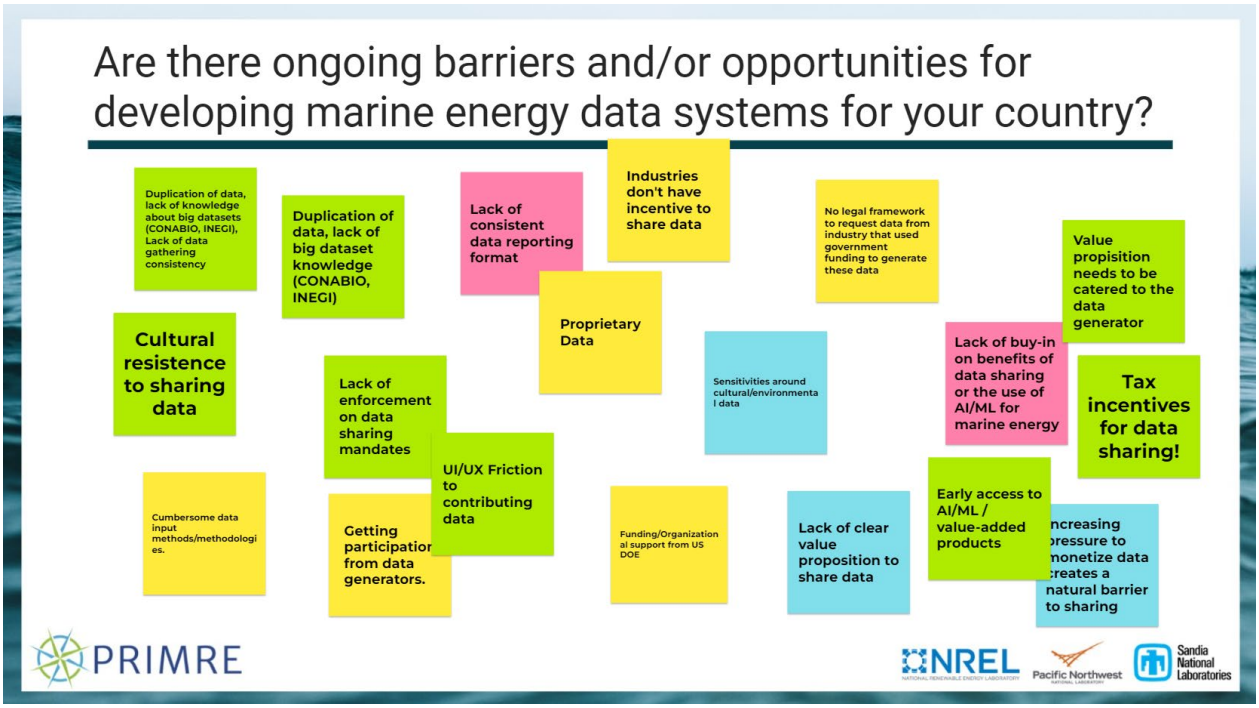
Are the appropriate permissions in place? Who is responsible?

PRIMRE

NREL Pacific Northwest Sandia National Laboratories



Breakout Group 2 (Jon)



Does your country mandate or encourage FAIR standards?

PRIMRE

NREL Pacific Northwest Sandia National Laboratories

What do you know about FARR? Is it something you will include in your systems?

PRIMRE

NREL Pacific Northwest Sandia National Laboratories

*Breakout group did not have enough time to complete the final slide.

Are you considering machine readability to your data?

yes

No

Yes!

PRIMRE

NREL Pacific Northwest Sandia National Laboratories

What issues are you having accessing data, archiving data, finding places to put data, on marine energy?

Access: Data doesn't exist or it is too low resolution.

pre-rev companies are holding onto every piece of information

I think we need to advertise the marine data pipeline more. They need to be presenting their product.

regulatory agencies data (US FERC E-Library)

Adopting new data standards creates interoperability issues with our own data, which has a higher impact to us than data sharing presently

Companies do not share their data, few government data, the datasets available are not well know

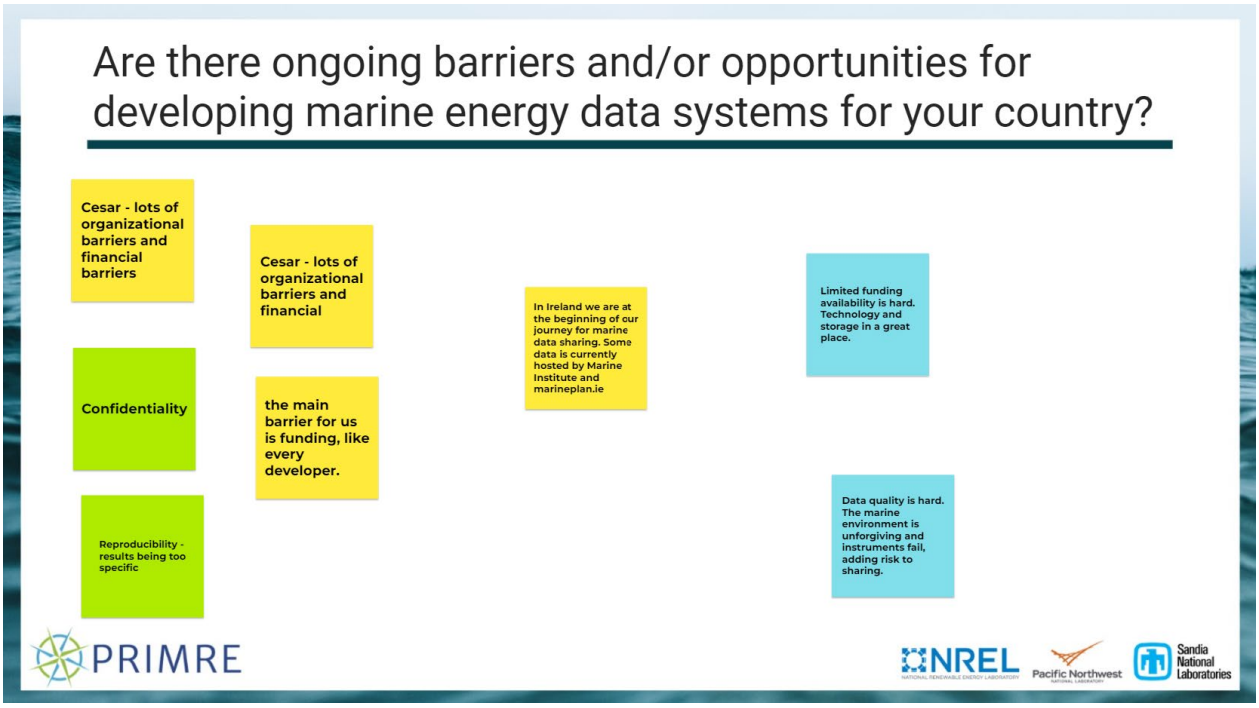
Data discoverability. I don't know what data is out there.

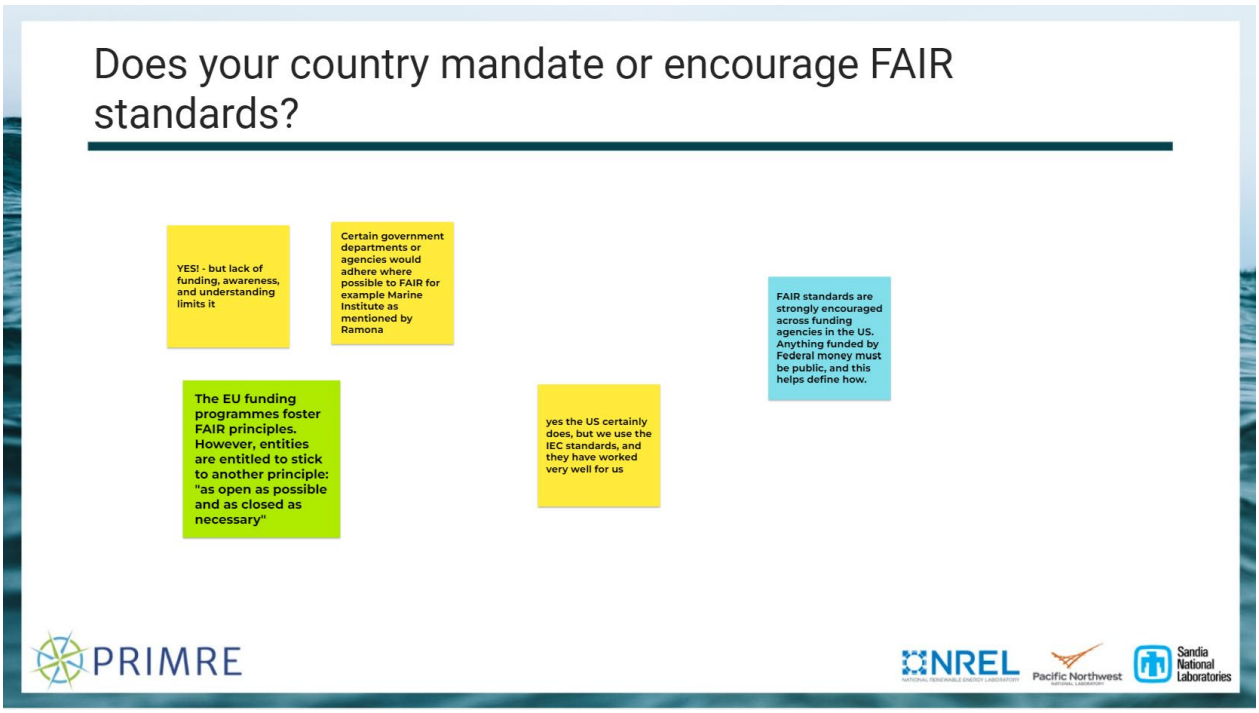
Data moratorium can prevent access to timely data and older data may not be as valuable (possibly more true for technology developers)

PRIMRE

NREL Pacific Northwest Sandia National Laboratories

Breakout Group 3 (Cesar/Jonathan)





What do you know about FARR? Is it something you will include in your systems?

Knew very little before knowing Jon Weers, but super interested in pushing this forward

I was not aware of it, but if makes data standardization more straightforward it sounds great

Walter, first time I heard about FARR

Preparing for AI!

PRIMRE

NREL Pacific Northwest Sandia National Laboratories

Are you considering machine readability to your data?

Where possible.

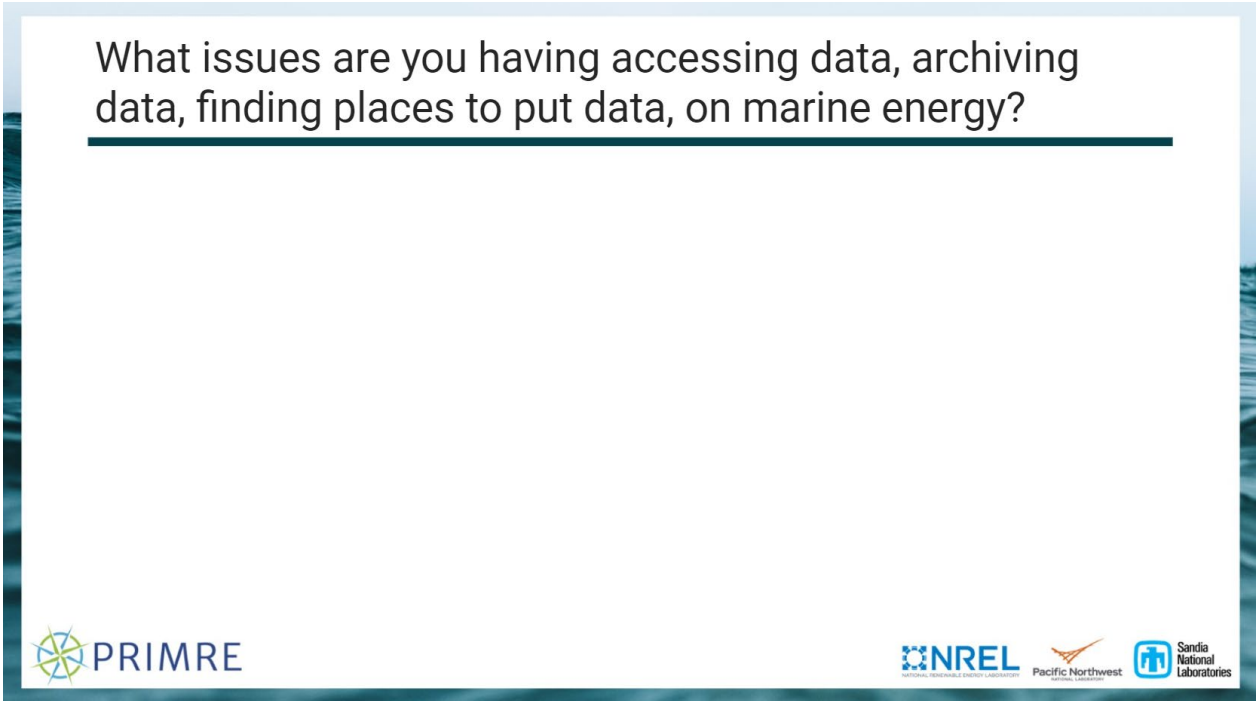
Structured data types (e.g., NetCDF and HDF5) help with this - pairs the data with metadata.

yes, certainly once we publish our data after commercialization

YES! Incorporating machine learning and AI into most of my computational work.

PRIMRE

NREL Pacific Northwest Sandia National Laboratories



*Breakout group did not have enough time to complete the final slide.