Enabling FAIR Data – The Importance of our Scientific Repositories

7 May 2020

Shelley Stall, AGU Sr. Director, Data Leadership
sstall@agu.org  @ShelleyStall
https://orcid.org/0000-0003-2926-8353
“Earth and space science data are a world heritage, and an essential part of the science ecosystem”

- From AGU’s Position Statement on Data
AGU’s position statement on data affirms that

“Earth and space science data are a world heritage, and an essential part of the science ecosystem. All players in the science ecosystem—researchers, repositories, publishers, funders, institutions, etc.—should work to ensure that relevant scientific evidence is processed, shared, and used ethically, and is available, preserved, documented, and fairly credited.”

https://www.agu.org/Share-and-Advocate/Share/Policymakers/Position-Statements/Position_Data
The AGU, along with our partners and over 300 stakeholders worldwide, convened a project – **Enabling FAIR Data** - that promoted the importance of data sharing (and software), citation, openness, and importance of data (and software) being FAIR as described in the FAIR Guiding Principles...
FAIR Guiding Principles

FAIR is...
Findable
Accessible
Interoperable
Reusable

FAIR Data Principles (applies to software and all digital objects)

• Findable
  – Assign persistent IDs (PIDs), provide rich metadata, register in a searchable resource, ...

• Accessible
  – retrievable by their ID using a standard protocol, metadata remain accessible even when data are no longer available...

• Interoperable
  – Use formal, broadly applicable languages, use standard vocabularies, qualified references...

• Reusable
  – Rich, accurate metadata, clear licenses, provenance, use of community standards...

Enabling FAIR Data Project in the Earth, Space, and Environmental Science
Repositories Have an Important Role

Provide the ability for datasets to be:
- Discovered
- Accessible
- Uniquely identified
- Well documented
- Reused with clear licensing
- Cited
- Machine readable (and human readable)
- Formatted for easy ingest into common tools
- Compliant with community vocabulary
- Linked to ORCIDs
- Linked to publications and other related research products
FAIR data relies on trusted repositories

...and the communities they bring together
Enabling FAIR Data asks Researchers to...

Locate trustworthy, community-accepted, FAIR-aligned repositories that support:

- **Documenting data and software** (and other research outputs as is possible) to agreed community standards that describe provenance and enable discovery, assessment of reliability, and reuse.
- **Persistent identifiers** for data and software (and other research outputs as is possible).
- **Licenses** for data and software (and other research outputs as is possible) that is as open as possible to enable the widest potential reuse.

Cite data, software, physical samples, and other research products.

Developing data availability statements.

Prepare and manage data management plans. Make them living documents.
Repository communities are key to supporting researchers to enable their data to be open and FAIR.

Without the necessary repository services supporting researchers, data can not be open and FAIR.
The Laura and John Arnold Foundation has awarded a grant to a coalition of groups representing the international Earth and space science community, convened by the American Geophysical Union (AGU), to develop standards that will connect researchers, publishers, and data repositories in the Earth, space, and environmental sciences to enable FAIR (findable, accessible, interoperable, and reusable) data on a large scale. This project will accelerate scientific discovery and enhance the integrity, transparency, and reproducibility of this data.

Commitment to Enabling FAIR Data in the Earth, Space, and Environmental Sciences

Publication of scholarly articles in the Earth, space, and environmental science community is conditional upon the concurrent
Enabling FAIR Data Project - Objectives

- **Data repositories** add value to research data, provide metadata and landing pages for discoverability, and support researchers with documentation guidance, citation support, and curation.

- **Earth, space, and environmental science publishers** align their policies to establish a similar experience for researchers. Data, software, technology will be available through citations that resolve to repository landing pages. Availability statements are provided.

**Data are NOT archived in the supplemental information of the paper.**

**Data ARE preserved in an appropriate trusted repository and cited in the paper.**
Enabling FAIR Data: Current Repository Signatories
(as of 7 May 2020)

Interdisciplinary Earth Data Alliance
Ubiquity Press
Biological and Chemical Oceanography Data Management Office
Paleobiology Database
California Digital Library – CDL
Geological Data Center, Scripps Institution of Oceanography
CCHDO (CLIVAR and Carbon Hydrographic Office)
Digital Rocks Portal
Environmental Data Initiative
PANGAEA, Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI), Center for Marine Environmental Sciences, University of Bremen (MARUM)
DIW Berlin
WDC Climate, Deutsches Klimarechenzentrum (DKRZ)
GFZ Data Services
Mendeley Data
Magnetics Information Consortium (MagIC)
Consortium of Universities for the Advancement of Hydrological Science, Inc (CUAHSI)

UK National Geoscience Data Centre (NGDC)
Network for Computational Modeling in the Social and Ecological Sciences
Figshare
4TU.Centre for Research Data
GigaScience
UNAVCO
OSGeo
Neotoma Paleoecology Database
OpenTopography
Dataverse
Edinburgh DataShare
UK National Geoscience Data Centre (NGDC)
Network for Computational Modeling in the Social and Ecological Sciences
Figshare
4TU.Centre for Research Data
GigaScience
UNAVCO
OSGeo
Neotoma Paleoecology Database
OpenTopography
Dataverse
Edinburgh DataShare
Neotoma Paleoecology Database
OpenTopography
Dataverse
Edinburgh DataShare
National Ecological Observatory Network (NEON)
EnviDat (www.envidat.ch)
Ocean Networks Canada, UVIC
Arctic Data Center
KNB Data Repository
Dryad
Federation University Australia
Data Repository for the University of Minnesota
Texas Digital Library

What Steps are Still Needed... for all Repositories to provide the support to Researchers to Enable FAIR Data?
Resources for this talk

- **AGU's Data Position Statement**: [https://www.agu.org/Share-and-Advocate/Share/Policymakers/Position-Statements/Position_Data](https://www.agu.org/Share-and-Advocate/Share/Policymakers/Position-Statements/Position_Data)
- **Enabling FAIR Data Project**: [https://copdess.org/enabling-fair-data-project/](https://copdess.org/enabling-fair-data-project/)
Thank you.

Shelley Stall, AGU Senior Director, Data Leadership
sstell@agu.org @ShelleyStall
https://orcid.org/0000-0003-2926-8353