Enabling FAIR Data in the Earth, Space, Environmental Sciences

Big Data Interagency Working Group

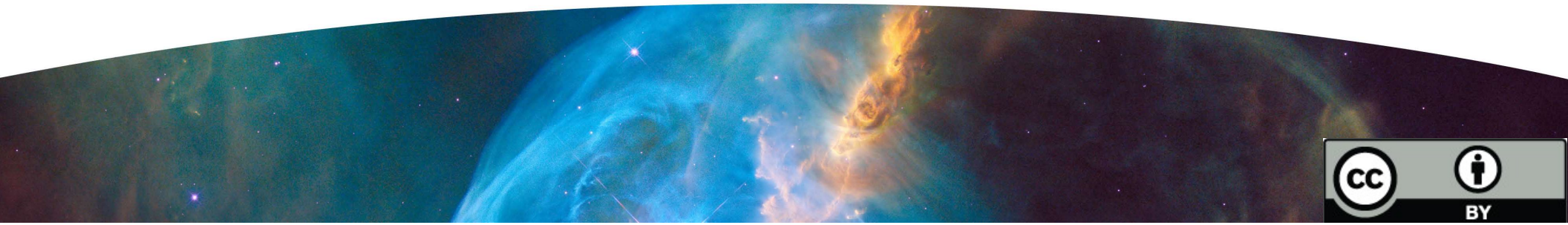
23 August 2018

Shelley Stall, AGU Director, Data Programs

sstall@agu.org @ShelleyStall

Project Steering Committee:

Lynn Yarmey (RDA/RPI)
Erin Robinson (ESIP)
Kerstin Lehnert (IEDA)
Mark Parsons (RPI / TWC)
Lesley Wyborn (NCI/ARDC/AuScope) Brooks Hanson (AGU)
Brian Nosek (Center for Open Science) Joel Cutcher-Gershenfeld (Brandeis)



American Geophysical Union

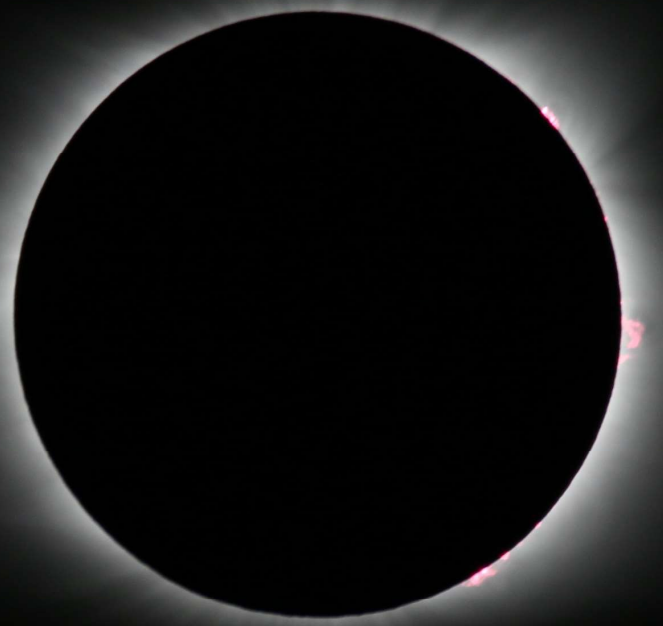
- > 60,000 members across 144 countries
- 20 peer-reviewed scholarly journals
- 100 year anniversary coming in 2019
- Scientific meetings
- Eos.org - online and print magazine
- ESSOAr – preprint server

Galvanizes a community of earth and space scientists that collaboratively advances and communicates science and its power to ensure a sustainable future.

AGU 100 ADVANCING EARTH AND SPACE SCIENCE

AGU's position statement on data affirms that

"Earth and space sciences data are a world heritage. Properly documented, credited, and preserved, they will help future scientists understand the Earth, planetary, and heliophysics systems."

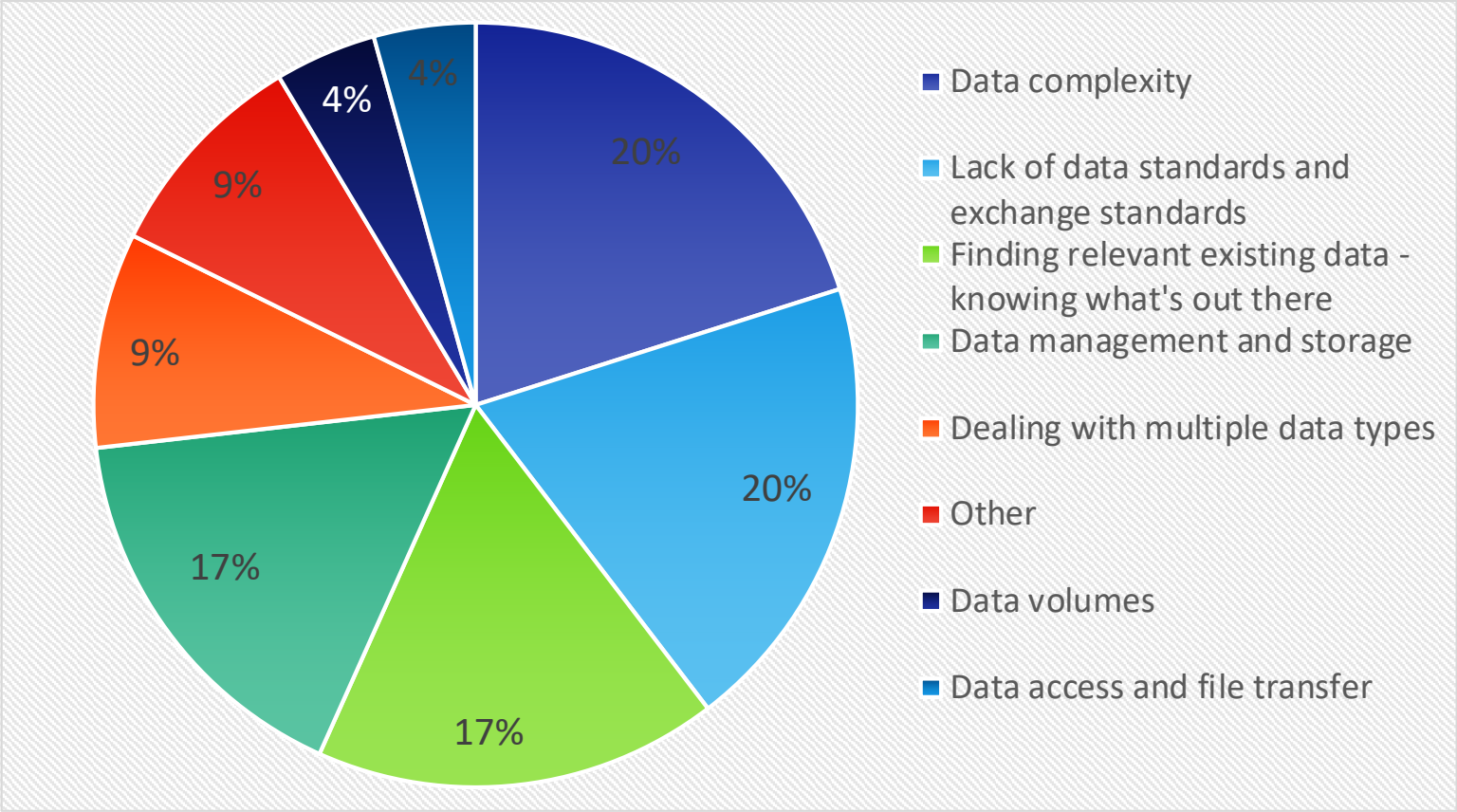


https://sciencepolicy.agu.org/files/2013/07/AGU Data Position Statement Final 2015.pdf

Photo Credit: Photo by Rick Meyers on Unsplash

Researcher Challenges with Data Use

The top four issues accounted for 73% of respondents



Data Management Skills Gap Analysis, April 7, 2017
http://www.bfe-inf.org/resource/skills-gap-analysis

BELMONT FORUM | e-INFRASTRUCTURES DATA MANAGEMENT

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data.

- From *The FAIR Guiding Principles for scientific data management and stewardship*

FAIR Guiding Principles

FAIR is…

- **Findable**
- **Accessible**
- **Interoperable**
- **Reusable**

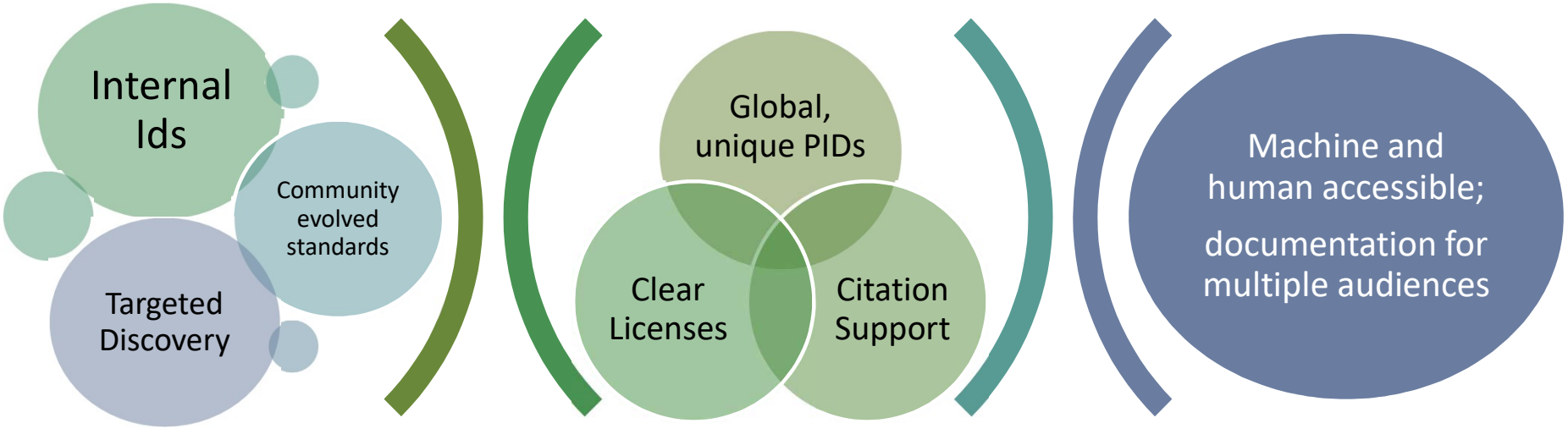
Article in Nature journal *Scientific Data*: Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

FAIR Data Principles (applies to software too)

- 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…

Article in Nature journal *Scientific Data*: Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

Evolving Researcher Support



New Grant from Laura and John Arnold Foundations (LJAF)

Align publishers and repositories in following best practices to enable FAIR and open data and to create workflows so that researchers will have a simplified, common experience when submitting their paper to any leading Earth and space science journal.

This will accelerate scientific discovery and enhance the integrity, transparency, and reproducibility of this data.

Enabling FAIR Data Project - Objectives

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

- **FAIR-aligned 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 placed in the supplemental information.

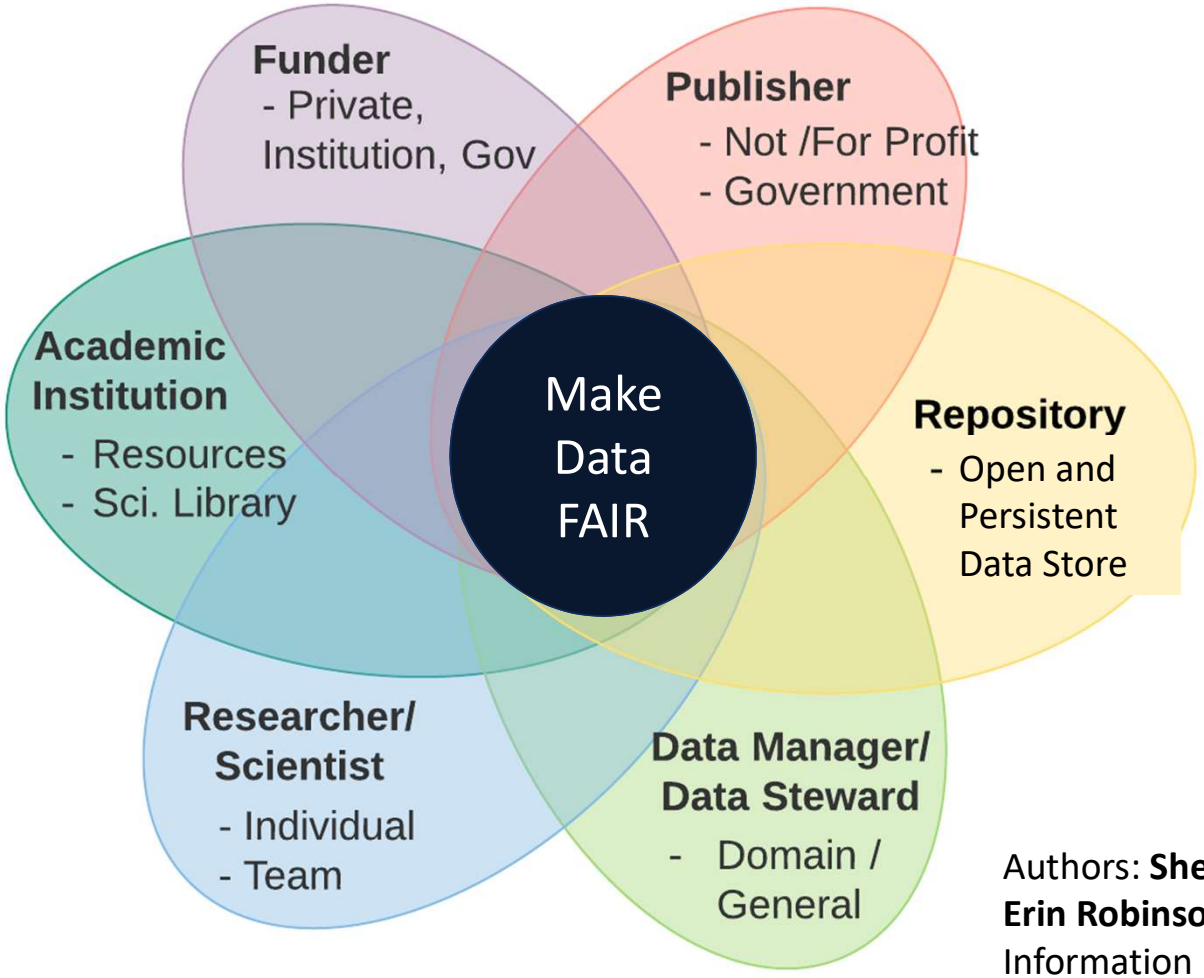
Community-Driven Project – Partnership Includes:

- **Science Data Communities**
  - AGU / EGU
  - Earth Science Information Partners (ESIP)
  - Research Data Alliance (RDA)
  - EarthCube / Council for Data Facilities
  - FORCE11
- **Publishers**
  - AGU
  - Proceedings of the National Academy of Sciences (PNAS)
  - Nature
  - Science/AAAS
  - Elsevier
  - PLOS
  - Hindawi
  - Copernicus/EGU
  - Wiley

- **International Repositories (300+)**
- National Computational Infrastructure (NCI)
- AuScope
- Australian National Data Service (ANDS)
- Center for Open Science
- DataCite / re3data
- ORCID
- CrossRef
- CHORUS
- Scholix
- OSGeo
- Pangaea
- DataONE

And Growing!!

Research Data Ecosystem

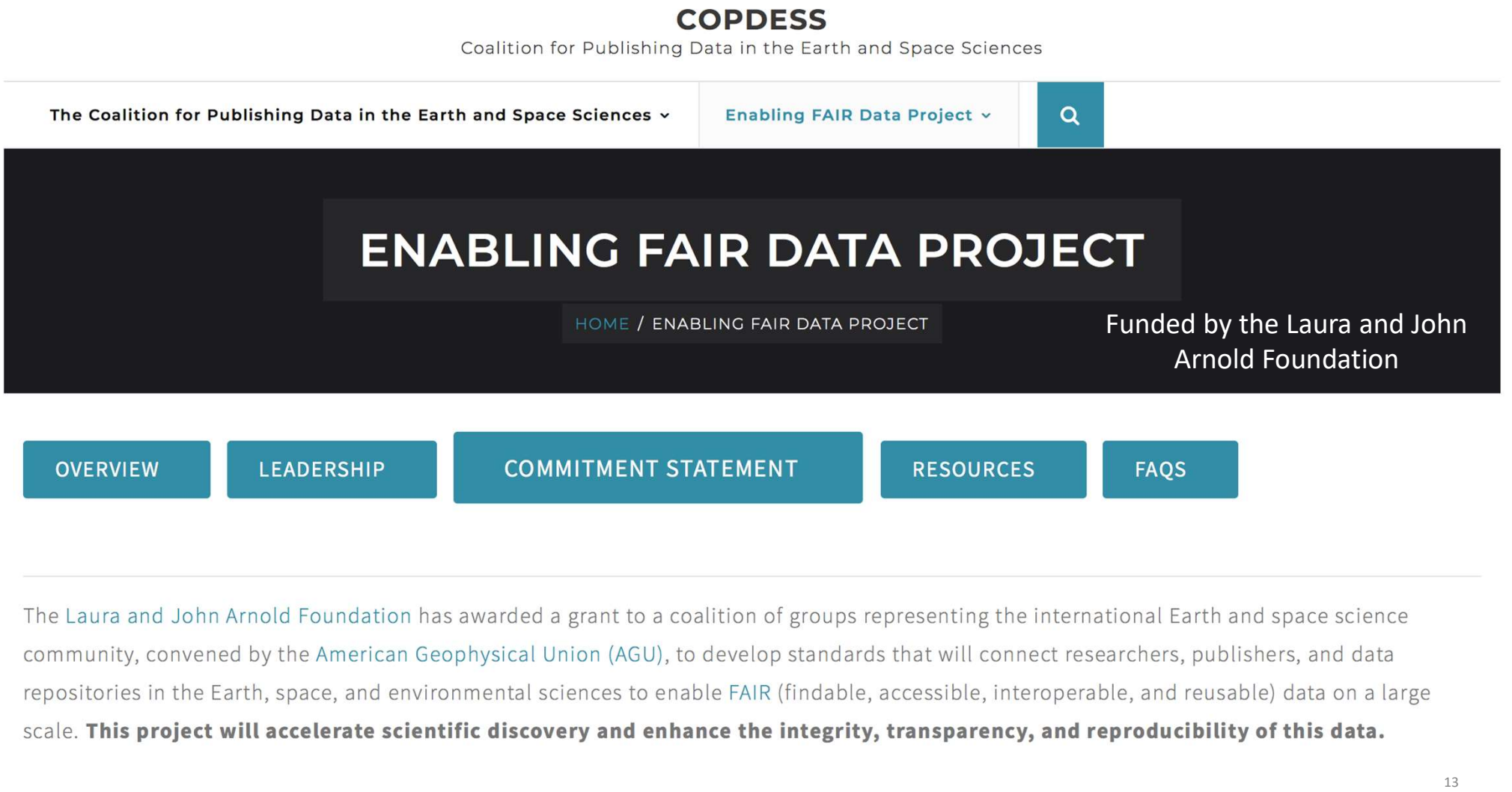


Other Roles:

- Research Labs
- Service providers to the ecosystem (e.g. PID providers like DataCite, github/Zenodo, CrossRef, CHORUS, Scholix)
- Research offices -- not at institutions (e.g. Ronin)
- International Efforts
- Societies
- Cyberinfrastructure
- IRBs

Authors: **Shelley Stall**, AGU Director of Data Program and **Erin Robinson**, Executive Director, Earth Science Information Partners

http://www.copdess.org/enabling-fair-data-project/



FAIR-aligned – Researcher Commitment

- Locating 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.
- Citing data, software, physical samples, and other research products
- Developing data availability statements
- Preparing and managing data management plans. Make them living documents.

FAIR-aligned: Repository Commitment

- Ensure that research outputs (e.g., data, software, technology, and physical samples) curated by repositories are open and FAIR, have essential documentation, and include human-readable and machine-readable metadata (e.g., on landing pages) in standard formats that are exposed and publicly discoverable.
- Ingest and expose data to promote interoperability and reuse.
- Ensure that unique, persistent identifiers are used for authors (e.g., ORCID), research objects (e.g., Digital Object Identifier), and physical samples (e.g., IGSN).
- Create associations among the research outputs that they manage and other related entities.
- Ensure that data and software have licenses that are as open as possible, and as protected as necessary.
- Support peer-review of related manuscripts by enabling access to the research outputs prior to publication.
- Gain third-party validation of trustworthy and sustainable practices and capabilities.

Project Next Steps – Invitations to Sign and Implement

1. **July 27th:** The Enabling FAIR Data “neutral” website ready to receive signatories of the commitment statement.
2. **August - September:** Engage with the first-line support to researchers: Editors, research offices, program officers, librarians, data managers, etc.,
3. **First week of September:** Completion of Repository Finder Tool. Developed by DataCite/re3data.
4. **Week of November 5th** - International Data Week: **Public announcement** of the Commitment Statement.
5. **Following the public release,** conduct researcher-focused webinar series - short 10-15mins episodes

Repository Finder Tool

Based on original work by Ruth Duerr, The Ronin Institute

Further designed by the Repository Guidance for Researchers (TAG A/D)
Co-Chairs: Danie Kinkade (BCO-DMO), Michael Witt (Purdue University)

Developed by: DataCite/re3data

Researchers: Completing two rounds of Usability testing

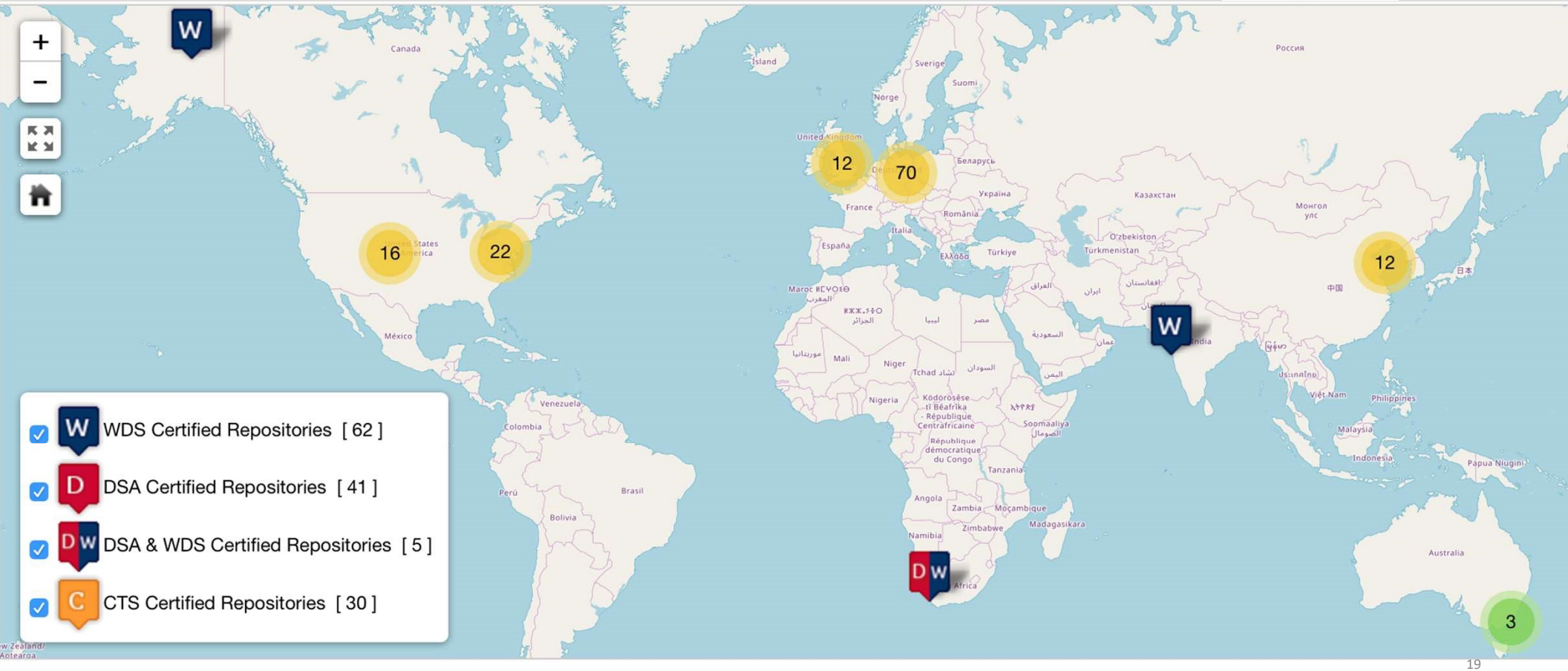
Repositories….

Repositories…Get Ready to be Found…

- Review your re3data.org record.
  - If you don't have one, go the page and select the "suggest" button at the top. Enter your repository information.
- **Tune your record** with this guide: http://bit.ly/RepoGuide
- Summarize your update in an email: info@re3data.org

CoreTrustSeal.org Certification

ICSU World Data System (WDS)



European Commission, Interim Report on FAIR Data & Action Plan

Please take the time to absorb these documents and provide your comments. The framework developed for how to move forward with FAIR is strong and viable -- valuable worldwide.

Turning FAIR data into reality: Interim report from the European Commission Expert Group on FAIR data
https://doi.org/10.5281/zenodo.1285271

FAIR Data Action Plan: Interim recommendations and actions from the European Commission Expert Group on FAIR data
https://doi.org/10.5281/zenodo.1285289

SCHOLIX: A Framework for Scholarly Link eXchange

- Establish links between scholarly literature and data.
- Understand what data underpins literature and what literature references data.
- Working with global aggregators of data-literature link information such as DataCite, CrossRef, OpenAIRE, or EMBL-EBI.
- These hubs in turn work with their natural communities of data centres or literature publishers to collect the information through existing community-specific workflows and standards.

ESIP Data Management Training EXPLORE

Home Browse Search Submit Help About Log in

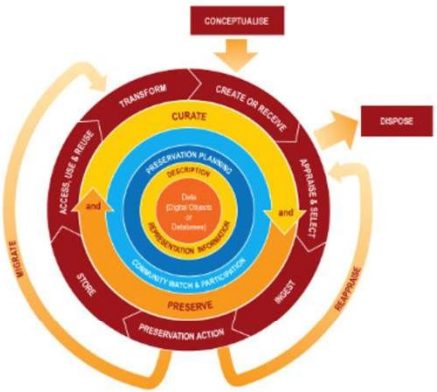
Welcome to the DMT Clearinghouse

The Data Management Training (DMT) Clearinghouse is a registry for online learning resources focusing on research data management.

It was created in a collaboration between the U.S. Geological Survey's Community for Data Integration, the Earth Sciences Information Partnership (ESIP), and DataONE.

For questions or feedback, please contact clearinghouseEd@esipfed.org

Read More



DCC Lifecycle - http://www.dcc.ac.uk/resources/curation-lifecycle-model

http://dmtclearinghouse.esipfed.org

Thank you!

Shelley Stall
Director, AGU Data Program
sstall@agu.org
@ShelleyStall

"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov