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

Polar Data Planning Summit

21 May 2018

Shelley Stall
Director, AGU Data Programs
sstall@agu.org

Ruth Duerr
Ronin Institute for Independent Scholarship
ruth.duerr@ronininstitute.org

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

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

AGU's position statement on data affirms that

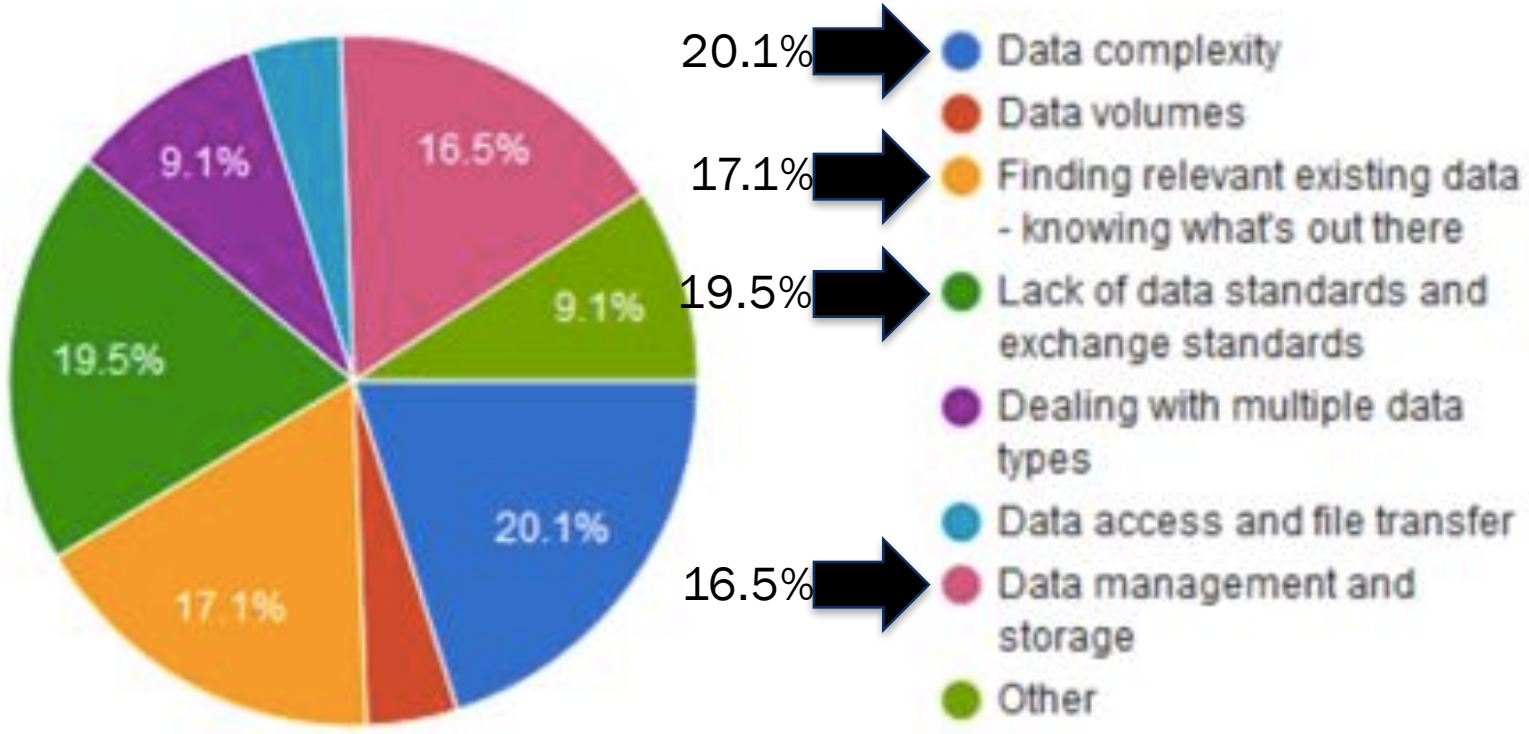
"Earth and space sciences data are a world heritage. Properly documented, credited, and preserved, they will help future scientists understand the Earth, planetary, and heliophysics systems."

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://bfe-inf.org/document/skills-gap-analysis

BELMONT FORUM | e-INFRASTRUCTURES DATA MANAGEMENT

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).

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

- **FAIR-compliant Earth and space science publishers** will align their policies to establish a similar experience for researchers. Data will be available through citations that resolve to repository landing pages. Data are not placed in the supplement.

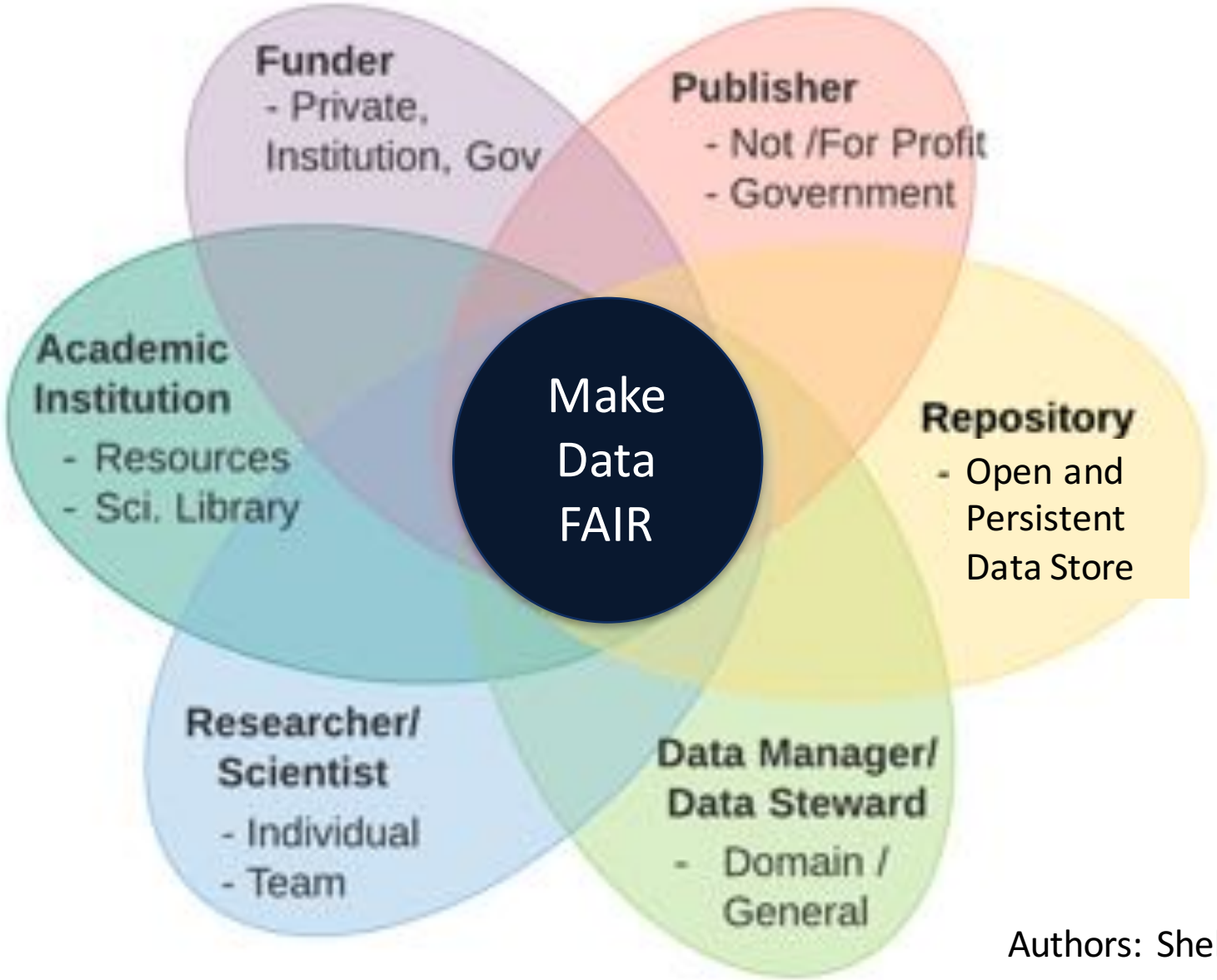
FAIR-Compliant Repositories

Services Provided:	Benefits:
Metadata support: Repository, Datasets, Citation	Supports Discovery, Understanding, Reuse
Persistent identifiers	Supports Data Citation and Credit for Data and Reuse
Citation / Landing page compliance	Supports Best Practices and Common Experience for Researchers
Publication Peer Review Support	Supports access for publication peer reviewers even if data isn't public yet.
Licensing policies (data and software)	Supports reuse of data and software.
Common list of approved FAIR-compliant repositories	Supports researchers locating compliant repositories.
Third-party validation of repository capability	Certification validating that many of the elements described above are implemented correctly in the repository (e.g. CoreTrustSeal, DataOne Member Node).

FAIR-Compliant Journals

Services Provided:	Benefits:
Ensure that data and other research outputs supporting publications are open and FAIR, and, to the greatest extent possible, **openly accessible at the time of publication**.	Supports FAIR-compliant repositories that are signatories of the Enabling FAIR Data effort.
Direct all core research outputs (data, software, appropriate samples and sample descriptions) to **trusted repositories**, following the FAIR Principles (e.g., using CoreTrustSeal certification). This means that supplements will no longer be used as the primary archive for data.	Supports FAIR-compliant repositories that are signatories of the Enabling FAIR Data effort. Data are no longer placed in the Supplemental Information. Improves research credit for data and reuse.
Adopt **a shared set of author instructions** that support these principles, providing a common set of expectations for authors in the Earth, space, and environmental sciences.	Supports a common experience for all researchers no matter what journal they select. Requirements for data citation, and data availability statements along with editor and reviewer workshops and training.
Provide **common expectations for publication peer review** when evaluating science and determining if the data and metadata are adequate.	Supports publication peer review process. Identifies what data and software should be available to editors and peer reviewers when evaluating the science.
Implement common workflows for data citations	Supports best practices and providing a common experience for all researchers.

Research Data Ecosystem

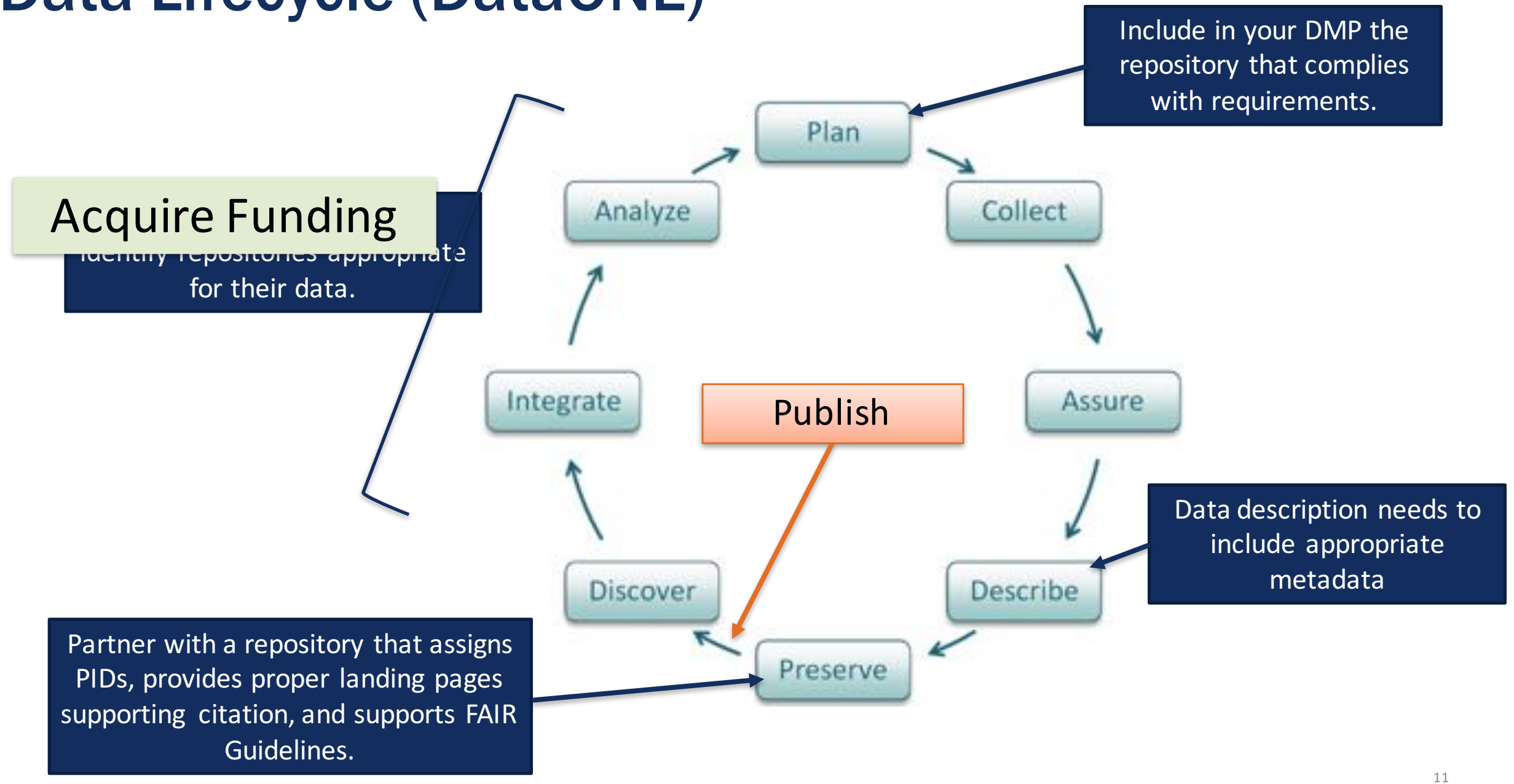


Other Roles:

- Research Labs
- Service providers to the ecosystem (e.g. PID providers like DataCite, github/Zenodo, CrossRef, FundRef, Scholix)
- Research offices -- not at institutions (e.g. Ronin)

Authors: Shelley Stall and Erin Robinson

Data Lifecycle (DataONE)



Timeline – 18 Months (@ 10 months remaining)

Preparation for 1st Stakeholder Meeting (Complete)	Aug 1, 2017 – Nov 15, 2017
First Stakeholder Meeting (Complete)	Nov 16 – 17, 2017
Working Groups Formed and Active	Nov 17, 2017 – Apr 2018
Development of Guidelines, Recommendations, and Policies for Journals and Repositories	Nov 17, 2017 – Apr 2018
Policies	Apr 2018 – July 2018
Second Stakeholder Meeting	September 12-13, 2018
Adoption and Implementation of Guidelines, Recommendations, and Policies Begins	Late Summer 2018

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

And Growing!!

Take Aways…

- Community-driven solution with AGU as convener
- Builds on the work previously done by COPDESS.org
- Data associated with publication will be open “by default”
- Quality of data documentation (metadata) becomes consistent – supports FAIR principles
- ESS Publishers and Repositories adopt project recommendations and guidelines

Enabling FAIR Data – Project Orientation Material

Article describing the Enabling FAIR Data Project:
https://eos.org/editors-vox/enabling-findable-accessible-interoperable-and-reusable-data

Outcome of the initial Stakeholder Meeting from Nov 16-17, 2017:
https://eos.org/agu-news/enabling-fair-data-across-the-earth-and-space-sciences

DataONE webinar recording:
https://www.dataone.org/webinars/enabling-fair-data

Join the Email List (hosted by RDA):
https://www.rd-alliance.org/groups/esipagurda-enabling-fair-data-coordination-group

Commitment Statement

(as of 17 April 2018)

Link to the Commitment Statement: http://bit.ly/FAIRCommitment

Comments period coming to a close soon.

Repositories Will Strive To:

- Ensure that **research outputs** they curate, including data, software, and samples are open and FAIR compliant and have essential and machine-readable metadata on landing pages in standard formats that are exposed and publicly discoverable.
- Ensure that these research outputs associated with publications have unique and **persistent identifiers** that support data citation following community practices.
- Include persistent unique identifiers and standard information for **samples** (e.g. IGSN's that resolve to sample landing pages with comprehensive metadata).
- Support **publication peer-review** by enabling access to the research outputs prior to publication.

Publishers Will Strive To:

- Ensure that data and other research outputs supporting publications are open and accessible at the time of publication.
- Direct all core research outputs (data, software, appropriate samples and sample descriptions) to trusted repositories, following the FAIR Principles (e.g., using CoreTrustSeal or Data Seal of Approval certification). This means that supplements will no longer be used as the primary archive for data.
- Adopt the following shared set of author instructions that support these principles, providing a common set of expectations for authors in the Earth, space, and environmental sciences. These include:
  - Adding data citations in the reference list following leading practices.
  - Including a "data availability statement" in published manuscripts that have research outputs (described here as the "TOP Statement"). This statement should be publicly accessible, machine readable and conform to common community standards.
  - Inform editors and reviewers of the FAIR data principles, and communicate these principles in author and reviewer workshops and training.
- Regularly review and update these data management practices, to align with current developments.
- Enable the connections between data citations in references and data, allowing researchers to receive credit for data sharing practices.

Thank You

Shelley Stall
Directory, AGU Data Programs
sstall@agu.org
@ShelleyStall