

Earth and Space Sciences: The Importance of Community Enabling FAIR and Open Data

#FORCE2017

October 26, 2017

Shelley Stall

AGU Director, Data Programs

ssall@agu.org

[@ShelleyStall](https://twitter.com/ShelleyStall)

The AGU logo consists of a stylized circular icon to the left of the letters "AGU". The icon is a white square with a smaller white square inside it, connected by a vertical line to a larger white square, all within a white-bordered box.

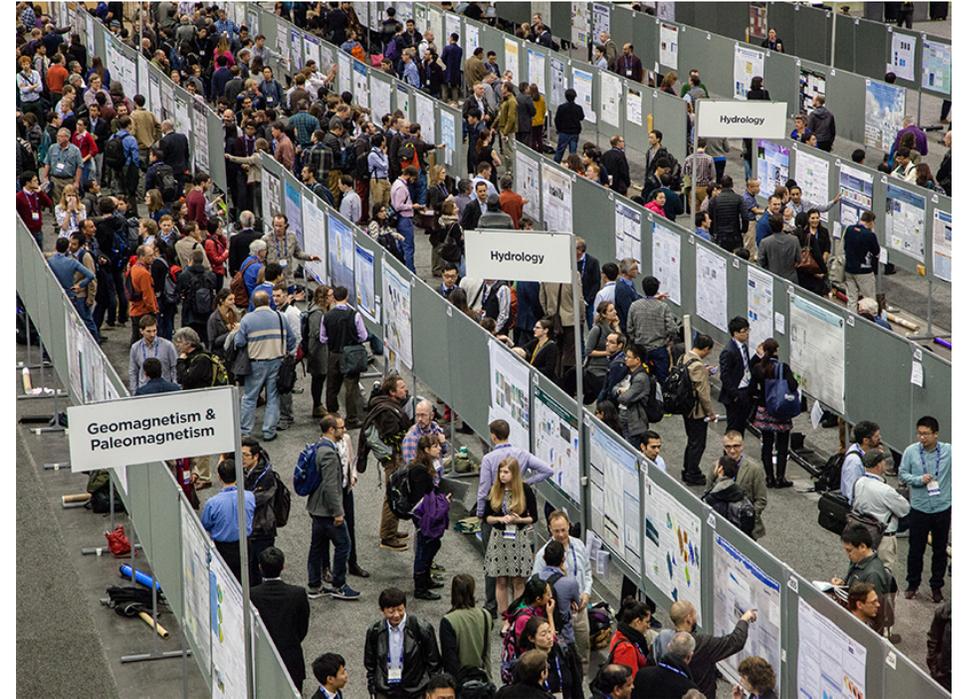
AGU

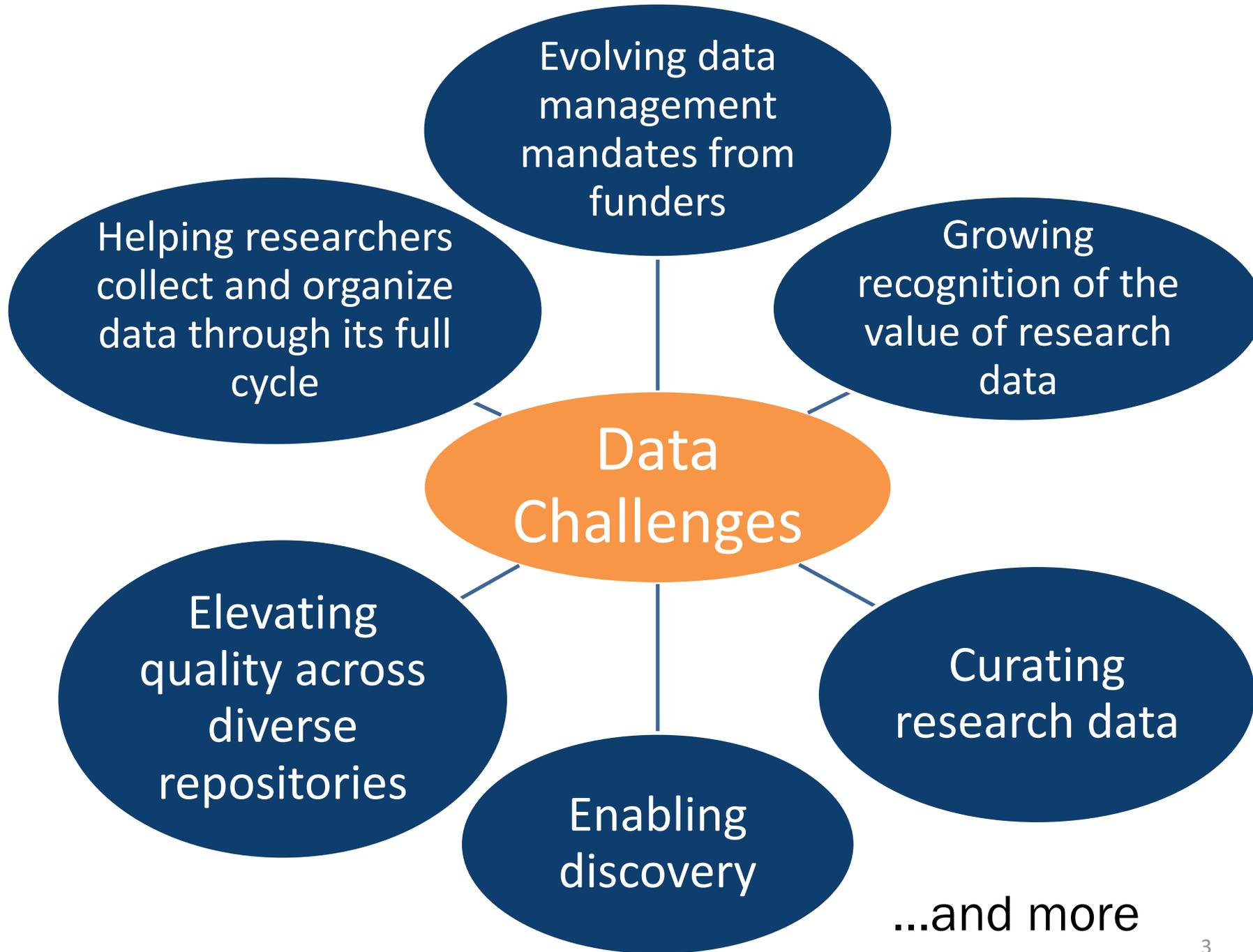


BY

About AGU

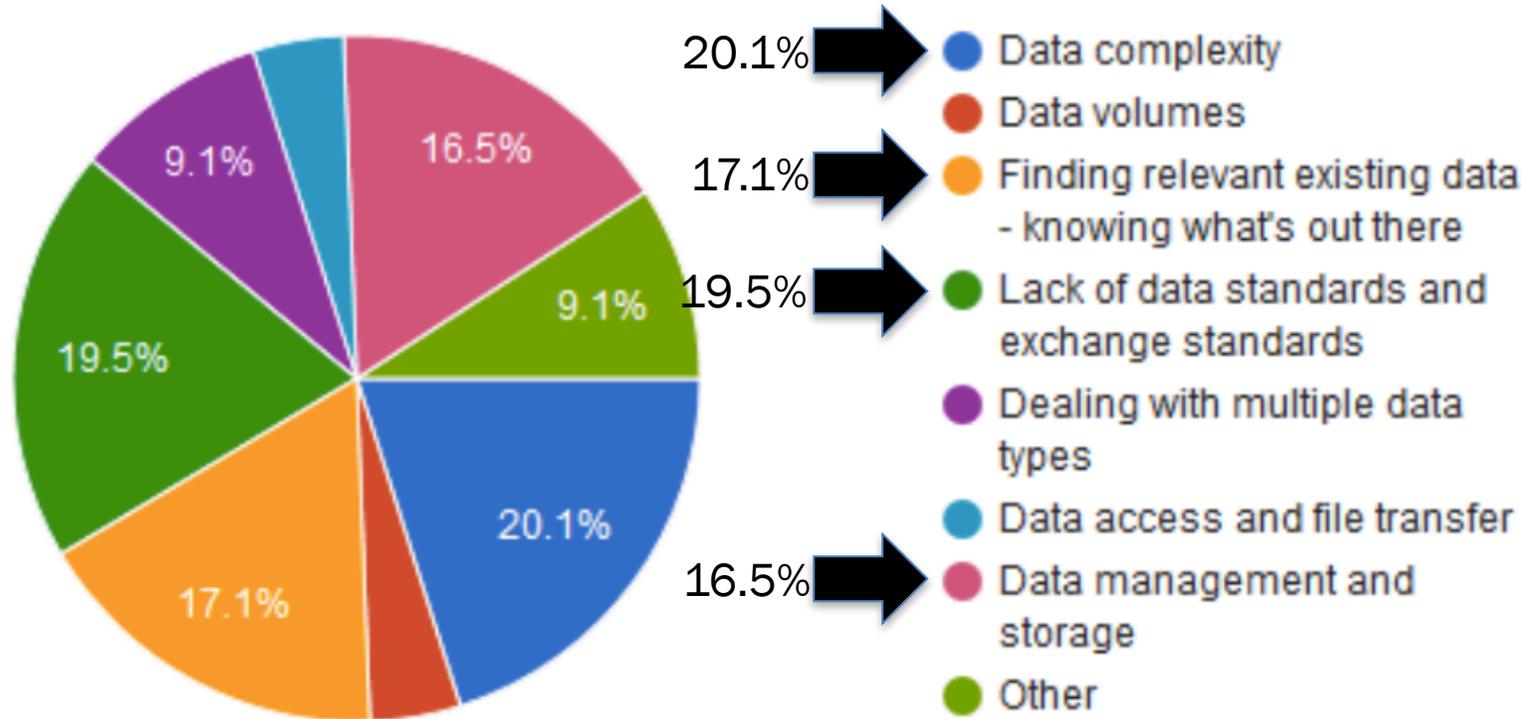
- Largest Earth and space science (ESS) society
 - 60,000 members; 137 countries
 - Much more than “geophysics”
 - Large annual meeting, 25,000 attendees
- Largest Society Publisher in the ESS
 - 20 peer-reviewed journals
 - >6000 published papers in 2016
- Outreach to government leaders and the public





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>

As a Researcher, you need...

Findable

- **Publications to include the data citation** that identifies the repository where the data is located. Use of persistent identifiers.
- Same for any relevant software.

Accessible

- **The repository to be accessible**, with clear information on the licensing of the data.

Interoperable

- **The data to be in a format appropriate** for the data domain. If multiple formats are used, the repository should have the ability to provide the data in all the accepted formats through conversion.

Reusable

- **The data in the repository to be well documented** – allows you to determine if it's “fit for use” without having to contact the PI.

FAIR Guiding Principles

Findable
Accessible
Interoperable
Reusable

Developed by Force11.org

Article in Nature: 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).



TRUE STORY – Dec 1, 2016

In the 3 June issue, *Science* published the Report “Environmentally relevant concentrations of microplastic particles influence larval fish ecology” by Oona M. Lönnstedt and Peter Eklöv (1). **The authors have notified *Science* of the theft of the computer on which the raw data for the paper were stored. These data were not backed up on any other device nor deposited in an appropriate repository.** *Science* is publishing this Editorial Expression of Concern to alert our readers to the fact that no further data can be made available, beyond those already presented in the paper and its supplement, to enable readers to understand, assess, reproduce, or extend the conclusions of the paper.

LETTERS

Edited by Jennifer Sills

Editorial expression of concern

IN THE 3 June issue, *Science* published the Report “Environmentally relevant concentrations of microplastic particles influence larval fish ecology” by Oona M. Lönnstedt and Peter Eklöv (1). The authors have notified *Science* of the theft of the computer on which the raw data for the paper were stored. These data were not backed up on any other device nor deposited in an appropriate repository. *Science* is publishing this Editorial Expression of Concern to alert our readers to the fact that no further data can be made available, beyond those already presented in the paper and its supplement, to enable readers to understand, assess, reproduce, or extend the conclusions of the paper.

Jeremy Berg
Editor in Chief

REFERENCE

1. O. M. Lönnstedt, P. Eklöv, *Science* **352**, 1213 (2016).

Published online 1 December
10.1126/science.aah6990

Retraction – May 3, 2017

Editorial Retraction

Jeremy Berg

+ See all authors and affiliations

Science 03 May 2017:
aan5763
DOI: 10.1126/science.aan5763

**absence of original data for the
experiments reported in the paper;**

University has not yet concluded its own investigation, the weight of evidence is that the paper should now be retracted. In light of the Board's recommendation and a 28 April 2017 request from the authors to retract the paper, *Science* is retracting the paper in full.



As a Repository supporting the Research Data Lifecycle...

Findable

- **Web accessible data services** that allow discovery tools to locate your data holdings.
- Persistent Identifiers (e.g. **Digital Object Identifiers – DOI**) for all data sets.
- Landing pages that support **data citations**.

Accessible

- **Tools** for researchers to access and investigate data sets that are of interest.

Interoperable

- **Metadata standards** and standard data set formats.

Reusable

- Use community recommended **domain metadata and vocabulary** to support transparency, decision making, and possible reuse.

Good Data Management – direct benefit to YOU the researcher

- Preserve the Scientific Record
- Enhancing your Reputation
- Reduce risk of data loss or quality
- Reduce need to repeat data gathering and be a good steward of public funds

Research Integrity, Transparency, Reproducibility

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.”



Data Management Assessment Program

AGU's DMMSM Best Practices and the Data Lifecycle

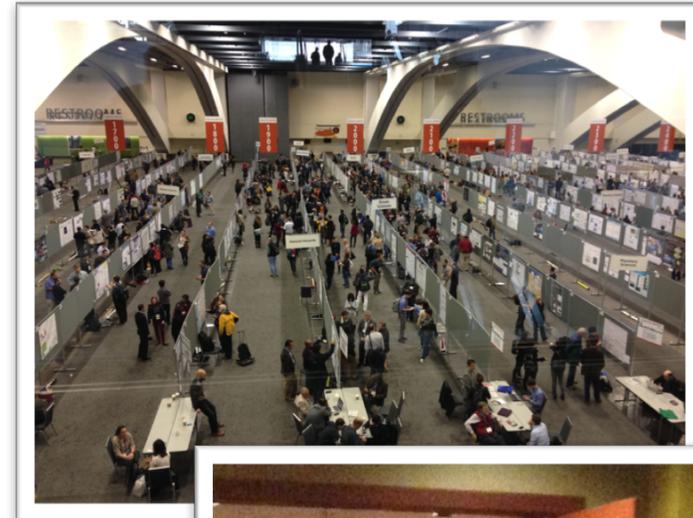


Practices Spanning Entire Lifecycle:

- Measurement & Analysis
- Process Management
- Process Quality Assurance
- Risk Management
- Configuration Management
- Sustainability, Consistency & Resilience

AGU Fall Meeting

- Over 25,000 scientists– oral presentations and posters
- Data Fair Key Note Speakers including DJ Patil, Rebecca Moore
- Data Fair – 3 Town Hall Panels





2016 AGU Fall Meeting: Data Fair

- Data Skills Recommended for Researchers and Scientists
- Data Management Plan Development and Management Best Practices
- Research Reproducibility Techniques

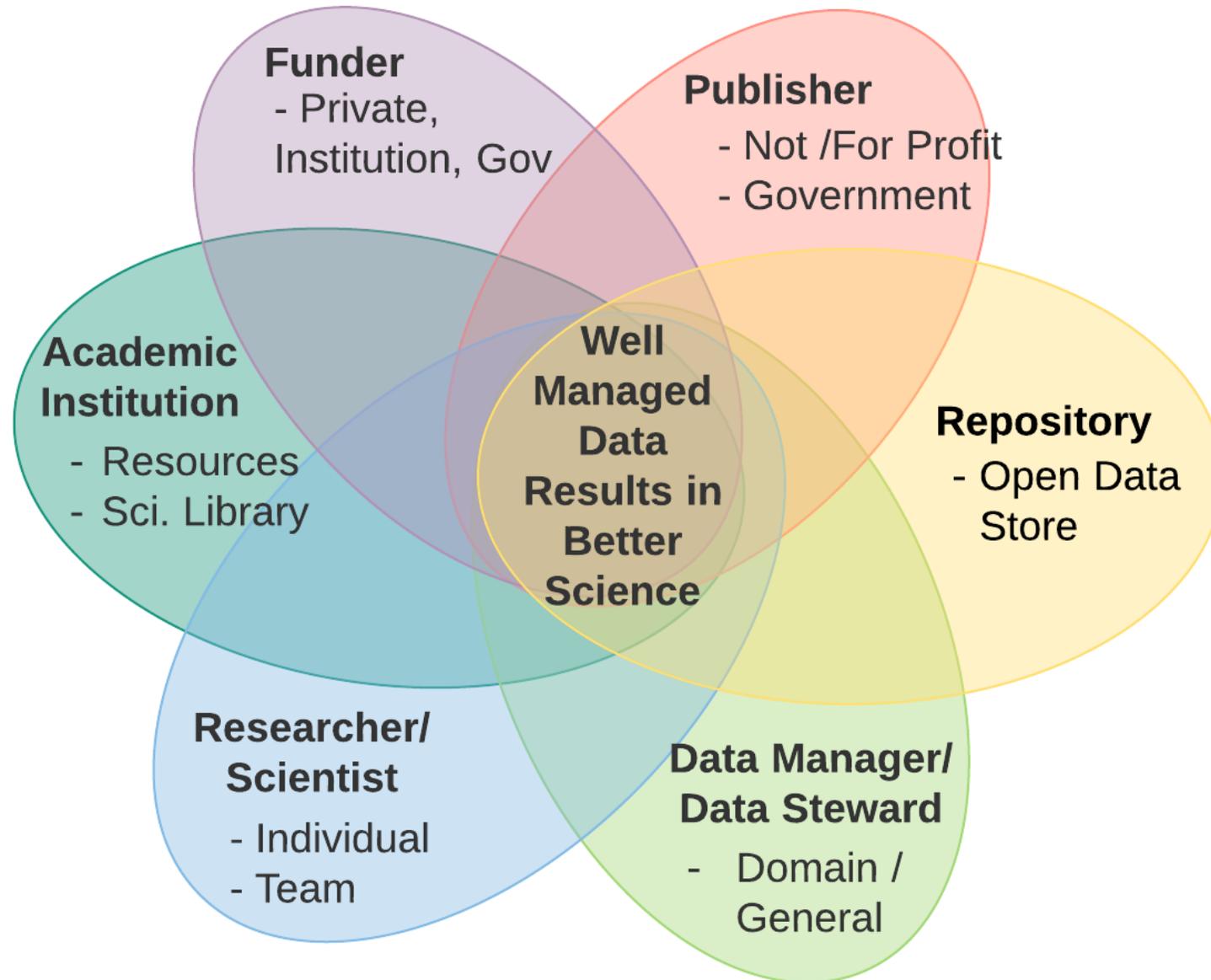


New at Fall Meeting: Data Help Desk



- Located in the Poster Hall
- Three Formats
 - Workshops
 - Demos
 - Data Reference Desk
- Sponsored by Earth and Space Science Informatics – AGU Focus Group
- Organized by ESIP

Research Data Ecosystem – Role View



Publishers and Repositories are Working Together...

- TOP (Transparency and Openness Promotion) guidelines, signed by 2900 journals and organizations
- COPDESS.org (Coalition on Publishing Data in the Earth and Space Sciences)—Statement of Commitment endorsed by most publishers and repositories in the Earth and space sciences
- Joint Declaration of Data Citation Principles endorsed by 114 organizations including most major publishers.
- Reproducibility conferences and outcomes (AAAS and other orgs)
- Quality/certification standards for repositories expanding

Challenge is practicing what you preach



Coalition on Publishing Data in the Earth and Space Sciences (COPDESS.org)

*Connecting Earth Science publishers and Data Facilities to help translate the aspirations of open, available, and useful data from **policy** into practice.*

- Formed in October 2014
- Endorsed a **Statement of Commitment, 2015**
- Includes: joint best practices between journals and repositories; references.

TOP – Modular Standards

Citation Standards Describes citation of data	Data Transparency Describes availability and sharing of data
Analytical Methods Transparency Describes analytical code accessibility	Research Materials Transparency Describes research materials accessibility
Design and Analysis Transparency Sets standards for research design disclosures	Preregistration of Studies Specification of study details before data collection
Preregistration of Analysis Plans Specification of analytical details before data collection	Replication Encourages publication of replication studies

The Problem...

At the time of publication:

Optional

- Are the data (and software/services) that support the paper properly documented and stored in a repository?

Optional

- Are the data citable with a persistent identifier, and support the FAIR Guidelines (Force11.org)?

No

- Do researchers have a similar experience with submitting their paper and supporting data (and software/services) no matter the journal?

The Solution... proper data documentation and storage



- In support of a publication...proper data documentation and storage in a repository.
 - Need to require data be included with the paper (as a citation and in the DAS) as the default option.
 - Need to engage repositories to ensure proper curation as much as possible.
 - Need to define what documentation (e.g. metadata) is essential and preferably optimal.

The Solution... proper data citation with a persistent identifier



- In support of a publication...proper data citation with a persistent identifier supporting the FAIR Guidelines.
 - Need to require use of repositories that use persistent identifiers.
 - Need to require use of repositories that have landing pages that support citation.
 - Need to require use of repositories that support FAIR Guidelines.

The Solution... provide a **similar experience for a researchers**



- In support of a publication...provide a **similar experience for a researcher** when submitting their paper and supporting data (and software/services) no matter the journal.
 - Journals and repositories need to define and adopt recommendations and align policies.



New Grant from Laura and John Arnold Foundations (LJAF)

Develop best practices and standards

that **will connect** researchers, publishers, and data repositories in the Earth and space sciences to enable **FAIR** data.

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

This project will help:

- 1) researchers understand and follow expectations regarding **data management and metadata required for publication**
- 2) publishers adopt and implement standard and best practices around **datasets, metadata, acceptable repositories, and citation supporting publication**
- 3) repository recognition of their valuable role in data lifecycle providing **curation services, persistent identifiers, and landing pages**

Community-Driven Project – Partnership Includes:

- **Science Data Communities**

- AGU
- Earth Science Information Partners (ESIP)
- Research Data Alliance (RDA)
- EarthCube / Council for Data Facilities

- **Publishers**

- AGU
- *Proceedings of the National Academy of Sciences (PNAS)*
- *Nature*
- *Science*

- **Repositories and COPDESS Signatories**

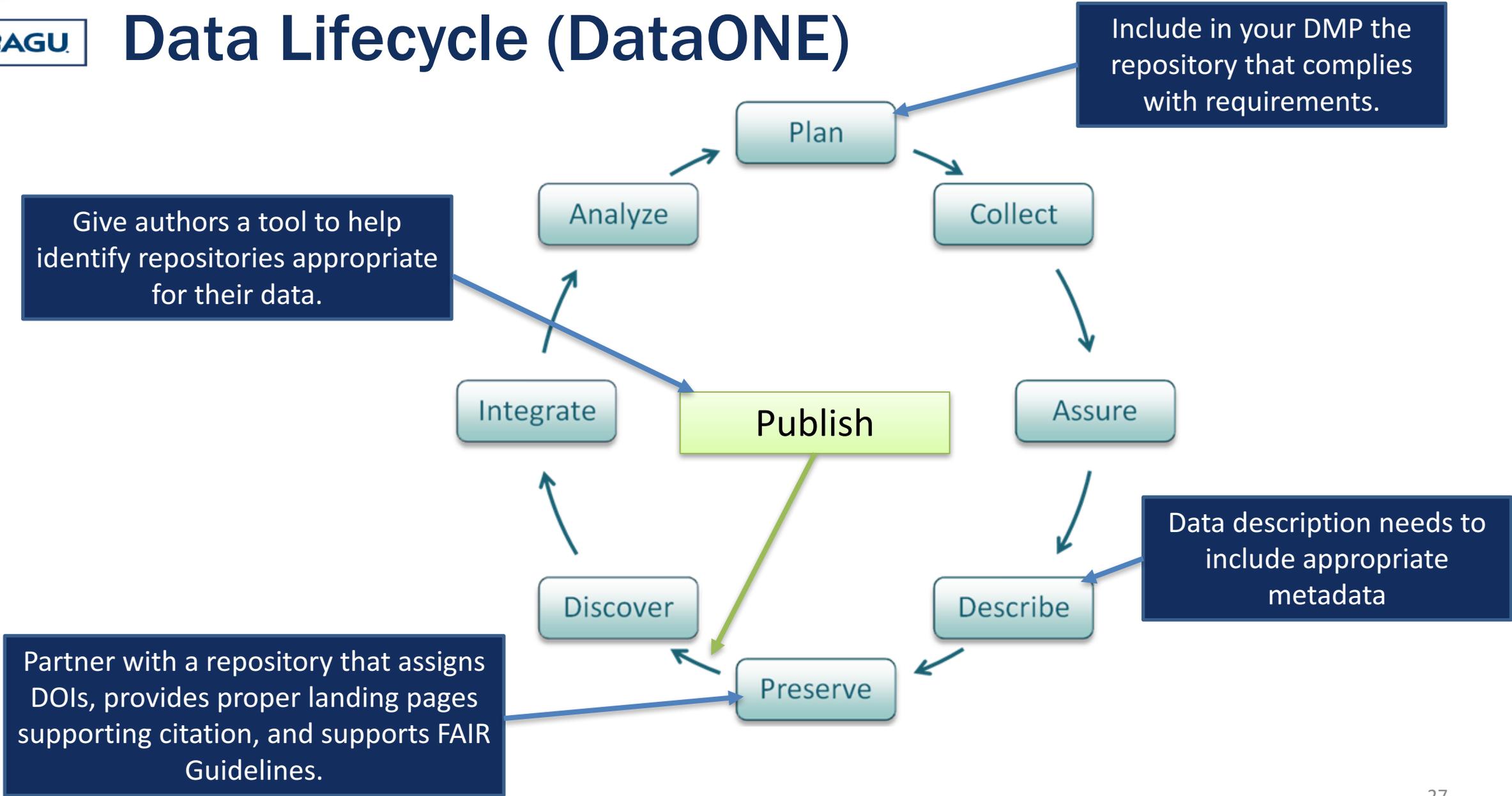
- National Computational Infrastructure (NCI)
- AuScope
- Australian National Data Service

- **Infrastructure**

- Center for Open Science

And Growing!!

Data Lifecycle (DataONE)



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

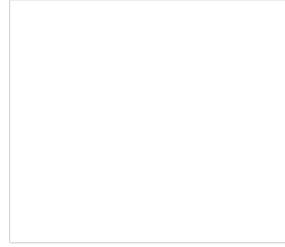
Timeline – 18 Months

Preparation for First Stakeholder Meeting	Aug 1, 2017 – Nov 15, 2017
First Stakeholder Meeting	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
- Testing of Guidelines, Recommendations, and Policies	Apr 2018 – July 2018
Second Stakeholder Meeting	July 2018
- Adoption and Implementation of Guidelines, Recommendations, and Policies Begins	July 2018

How To Participate...

- Stay Informed:
 - <http://www.copdess.org> -> Enabling FAIR Data Project
- Participate in the Stakeholder Alignment Survey – October 2017
- Participate in a Working Group
 - Formation is during First Stakeholder Meeting – Nov 2017
- Support FAIR Principals In the Rest of the Lifecycle
 - Incentives
 - Communication
 - Alignment

Questions?



Contact Information:

Shelley Stall

Director, AGU Data Programs,
Program Manager for Enabling FAIR Data
Project

@ShelleyStall

ssall@agu.org

