



Enabling FAIR Data

DataONE Webinar Series

September 12, 2017

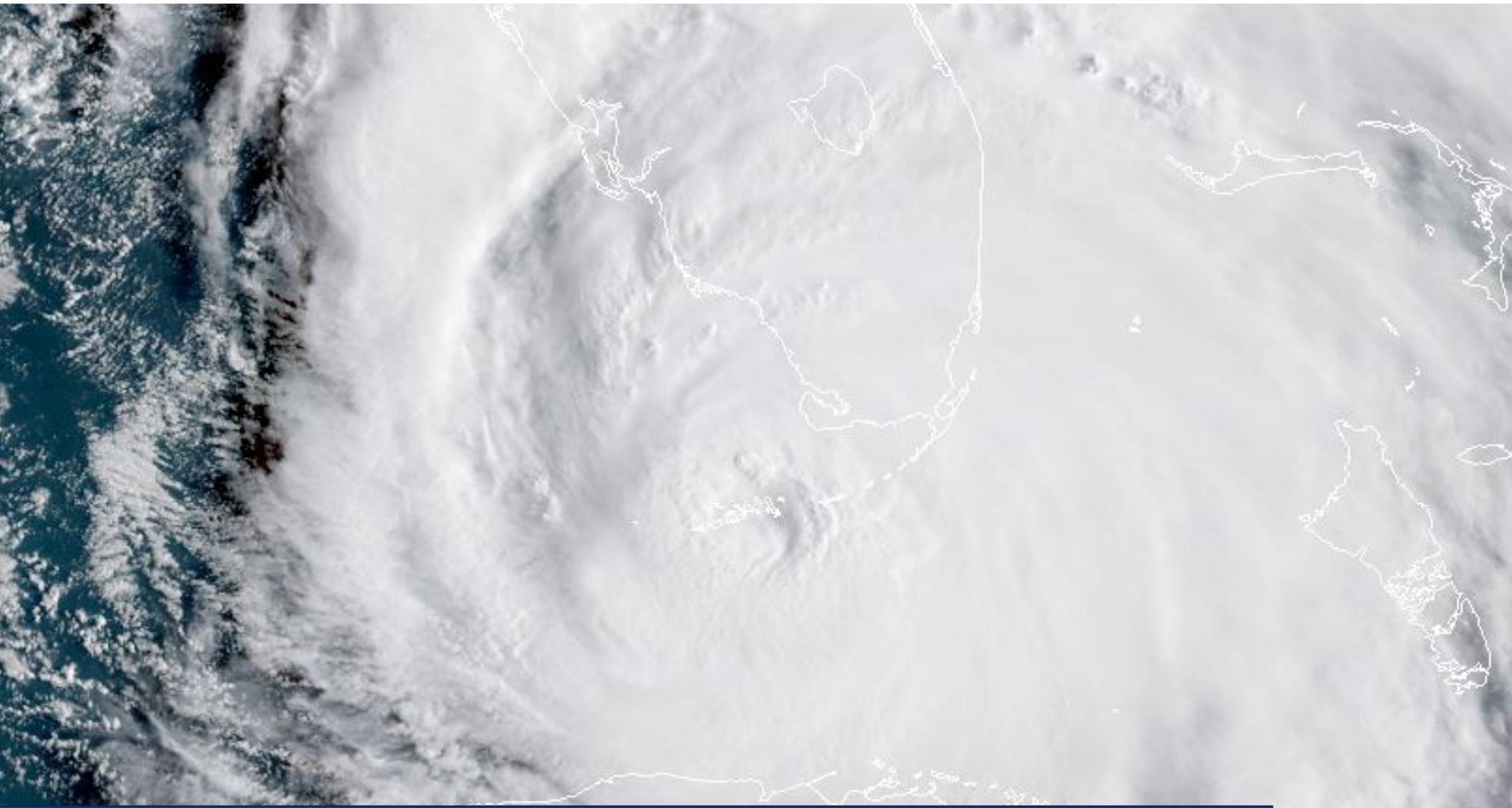
Shelley Stall

AGU Director, Data Programs

sstall@agu.org

@ShelleyStall

AGU



Satellites @NOAA Satellites Sep 10

S16 captured this geocolor image of #Irma's eye over the Florida Keys at x. 9:00 am EDT. Latest info @ <http://www.nhc.noaa.gov> .





American Geophysical Union

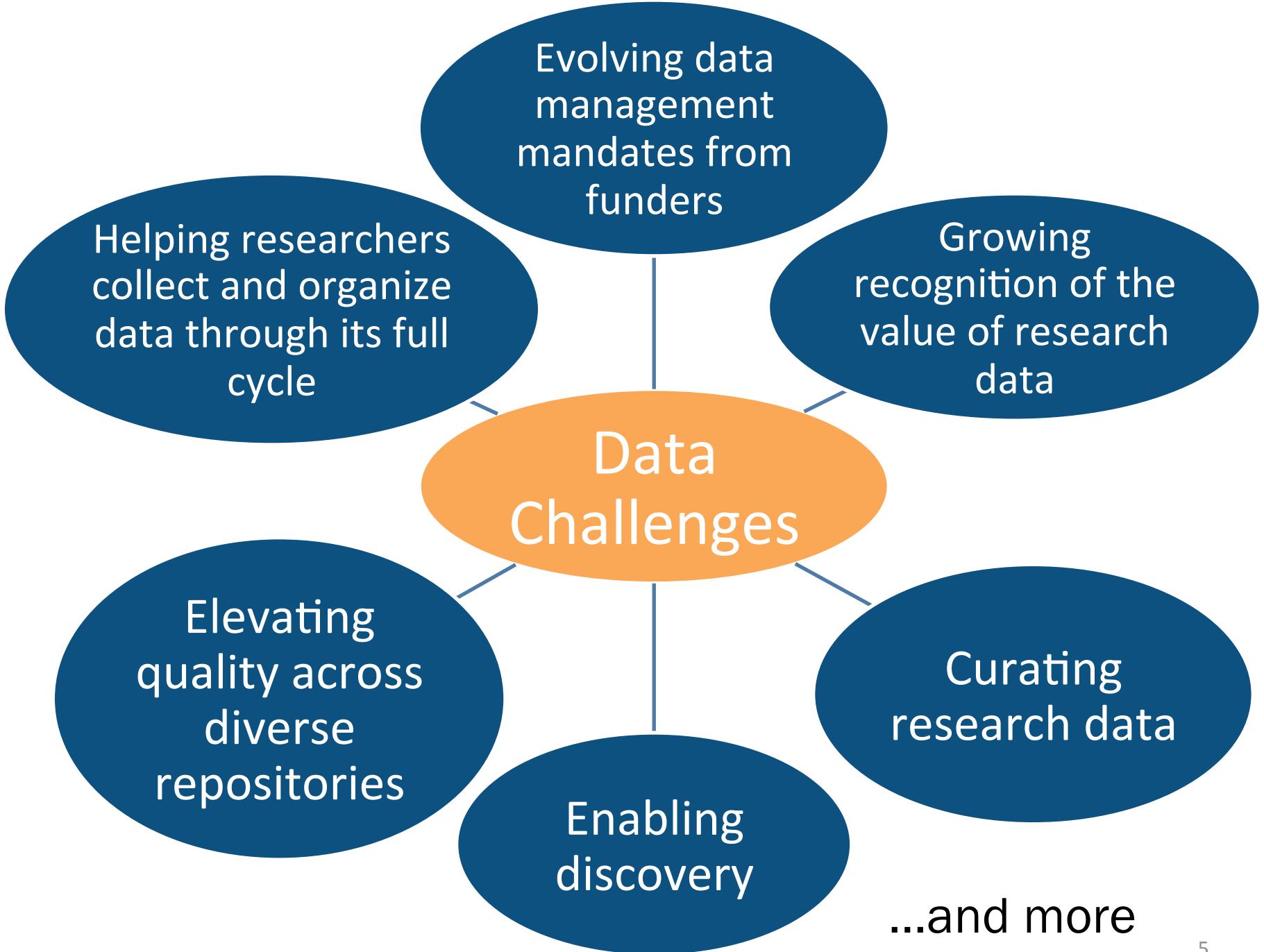
- > 60,000 members across 144 countries
- 20 peer-reviewed scholarly journals
- 100 year anniversary coming in 2019
- Scientific meetings
- Eos.org - online and print magazine

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



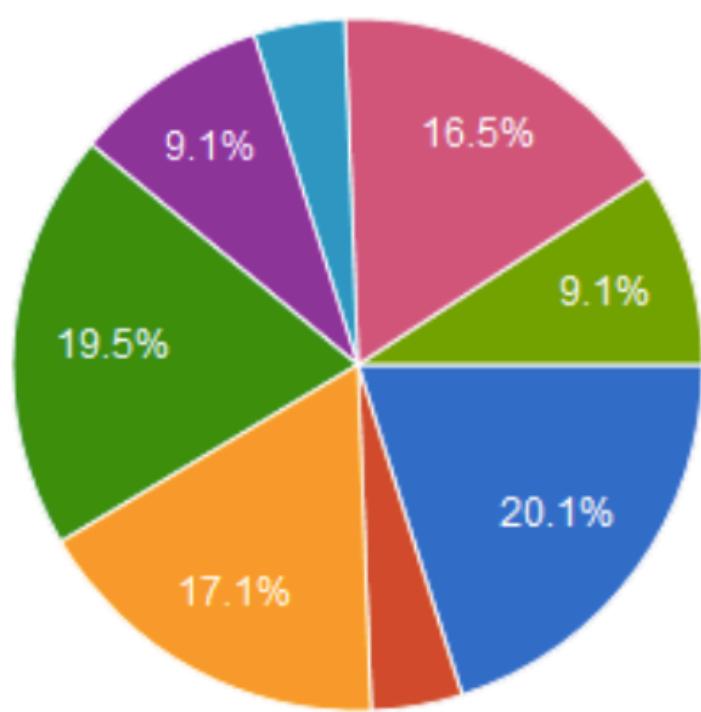
Late-Breaking Session on Hurricanes Harvey and Irma

- Abstract Submission Deadline: 31 October 2017, 11:59 ET
- <https://fallmeeting.agu.org>
- Conveners:
 - Guido Cervone (Lead), Associate Director of the Institute for CyberScience, Penn State
 - Jeffrey A. Nittrouer, Assistant Professor of Geology, Rice University, DE
 - Courtney Thompson, Assistant Professor, Department of Geography, Texas A&M University
 - Lesley Wyborn, Australian National University, Adjunct Fellow, National Computational Infrastructure Facility and Research School of Earth Sciences



Researcher Challenges with Data Use

The top four issues accounted for 73% of respondents



- ● Data complexity
- ● Data volumes
- ○ Finding relevant existing data
- knowing what's out there
- ● Lack of data standards and
exchange standards
- Dealing with multiple data
types
- ● Data access and file transfer
- ● Data management and
storage
- Other

Data Management Skills
Gap Analysis, April 7, 2017



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. Lonnstedt, P. Eklov, *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.

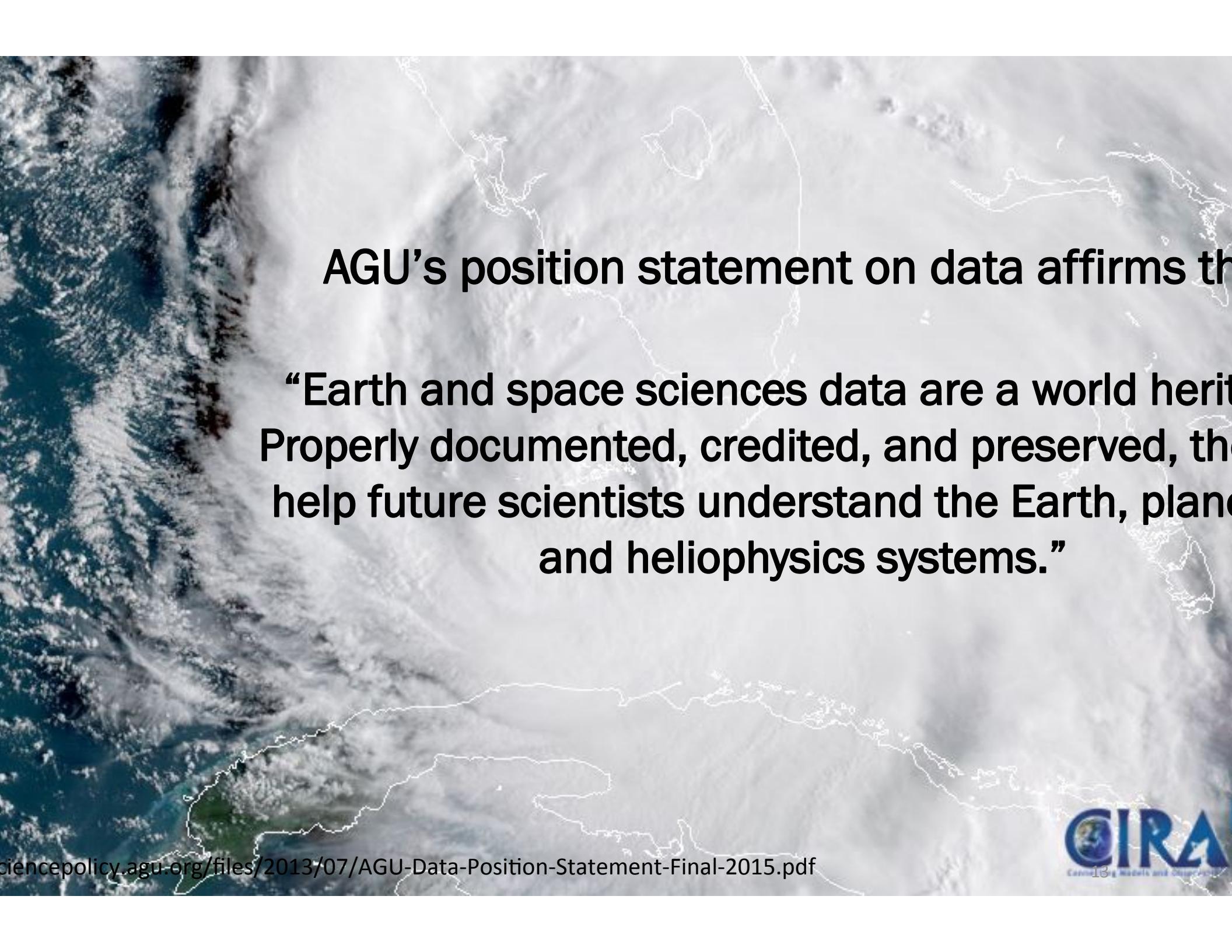
Reusable

- 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 the

“Earth and space sciences data are a world heritage.”
Properly documented, credited, and preserved, these data 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's Data Management Advisory Board

- **Chair**

Erin Robinson, Earth Science Information Partners

- **Members**

Tim Ahern, Incorporated Research Institutions for Seismology (IRIS)

Sky Bristol, USGS

Helen Glaves, British Geological Survey

Sophie Hou, National Center for Atmospheric Research /University Corporation for Atmospheric Research

Matt Jones, University of California Santa Barbara

M. Lynn Penn, Performance and Methods Consulting, LLC

H.K. Ramapriyan, Science Systems and Applications, Inc.

Jerry Sheehan, NIH/NLM

Victoria Stodden, University of Illinois at Urbana-Champaign

Carly Strasser, Moore Foundation

Lesley Wyborn, Australian National University

NJ Data Initiatives

- Data Management Training for Repository Operations
- Data Management Assessment Program

In Partnership -

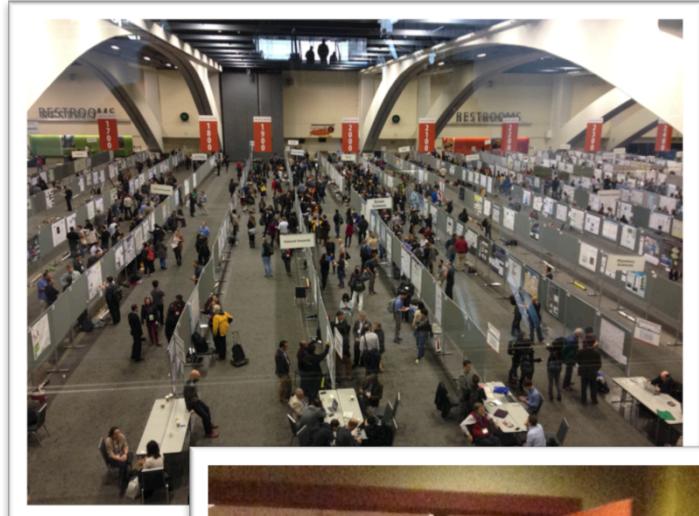


- Fall Meeting: Data Fair Key Notes, Researcher Engagement with ESIP
- Data Citation – Best Practices in Oral Presentation
- COPDESS.org – Coalition for Publishing Data in the Earth and Space Sciences with Publishers and Data Repositories
- Scientific Reproducibility with AAAS
- NASA's Software and Service Citation Working Group

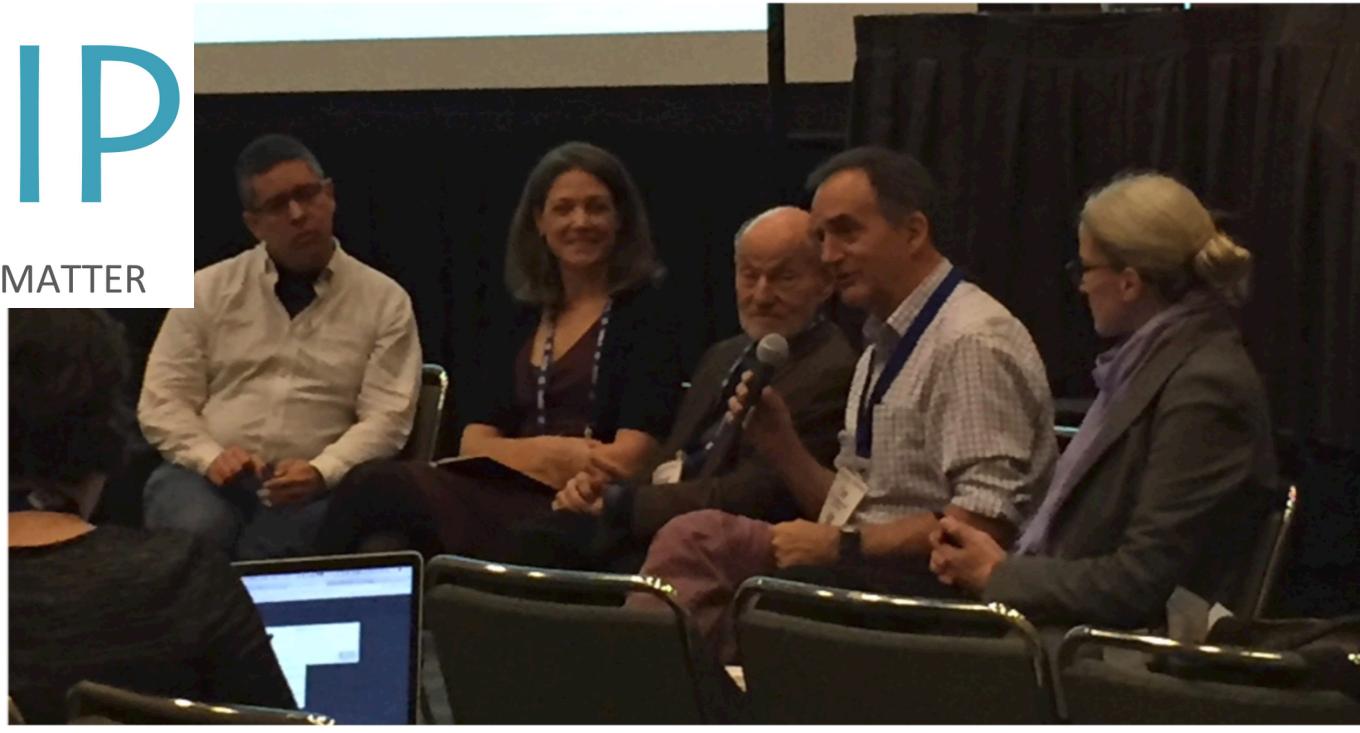


AGU Fall Meeting

- 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 Meet Data Fair



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



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



Why Include and Cite Data in your Oral Presentat

- Transparency and reproducibility in science requires sharing data and software.
- Support of good practice by including information on your data and software. Be the Role Model.
- Encourage reuse of your data by providing citation.

Recruiting your shoulders...

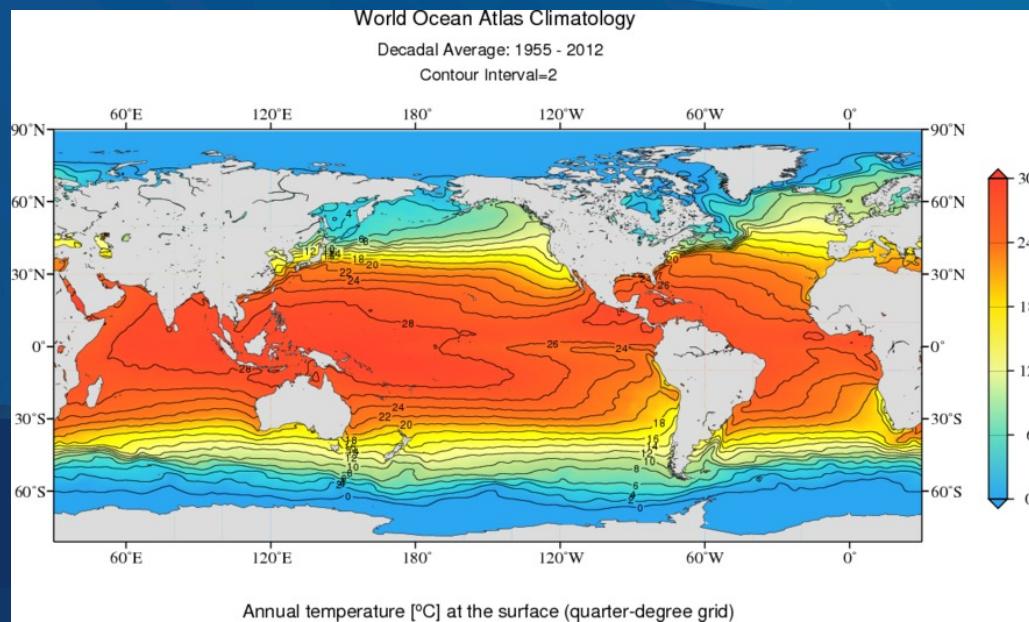


Photo credit: <https://i.ytimg.com/vi/ukSWNMxbygA/maxresdefault.jpg>

AGU 2016, IN53E-02

Based on NOAA's World Ocean Atlas 2013 v. 2

e.g., *Locarnini, R.A., A.V. Mishonov, J.I. Antonov, T.P. Boyer, H.E. Garcia, O.K. Baranova, and others. 2013. *World Ocean Atlas 2013 version 2 (WOA13 V2), Volume 1: Temperature*. In: NOAA National Centers for Environmental Information S. Levitus, ed, and A. Mishonov, technical ed, NOAA Atlas NESDIS 73, doi:10.7289/V55X26VD, www.nodc.noaa.gov/OC5/woa13/

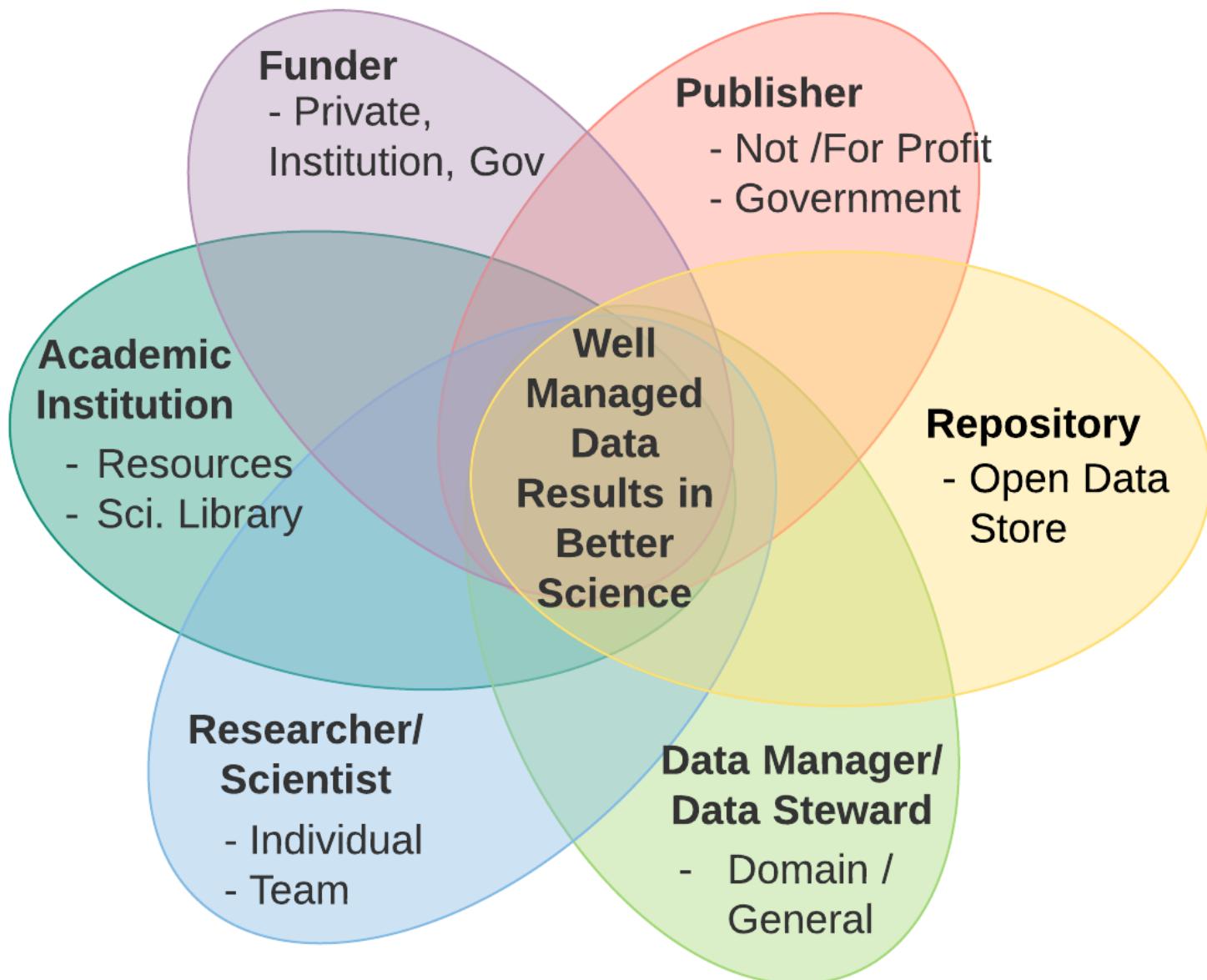


Temperature*
Salinity
Dissolved Oxygen

Nitrate
Silicate
Phosphate

Apparent Oxygen Utilization
Percent Oxygen Saturation

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 organizations)
- 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:

- Are the data (and software/services) that support the paper properly documented and stored in a repository?
- Are the data citable with a persistent identifier, and support the FAIR Guidelines (Force11.org)?
- Do researchers have a similar experience with submitting their paper 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 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 support citation.
 - Need to require use of repositories that support FAIR Guide



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**

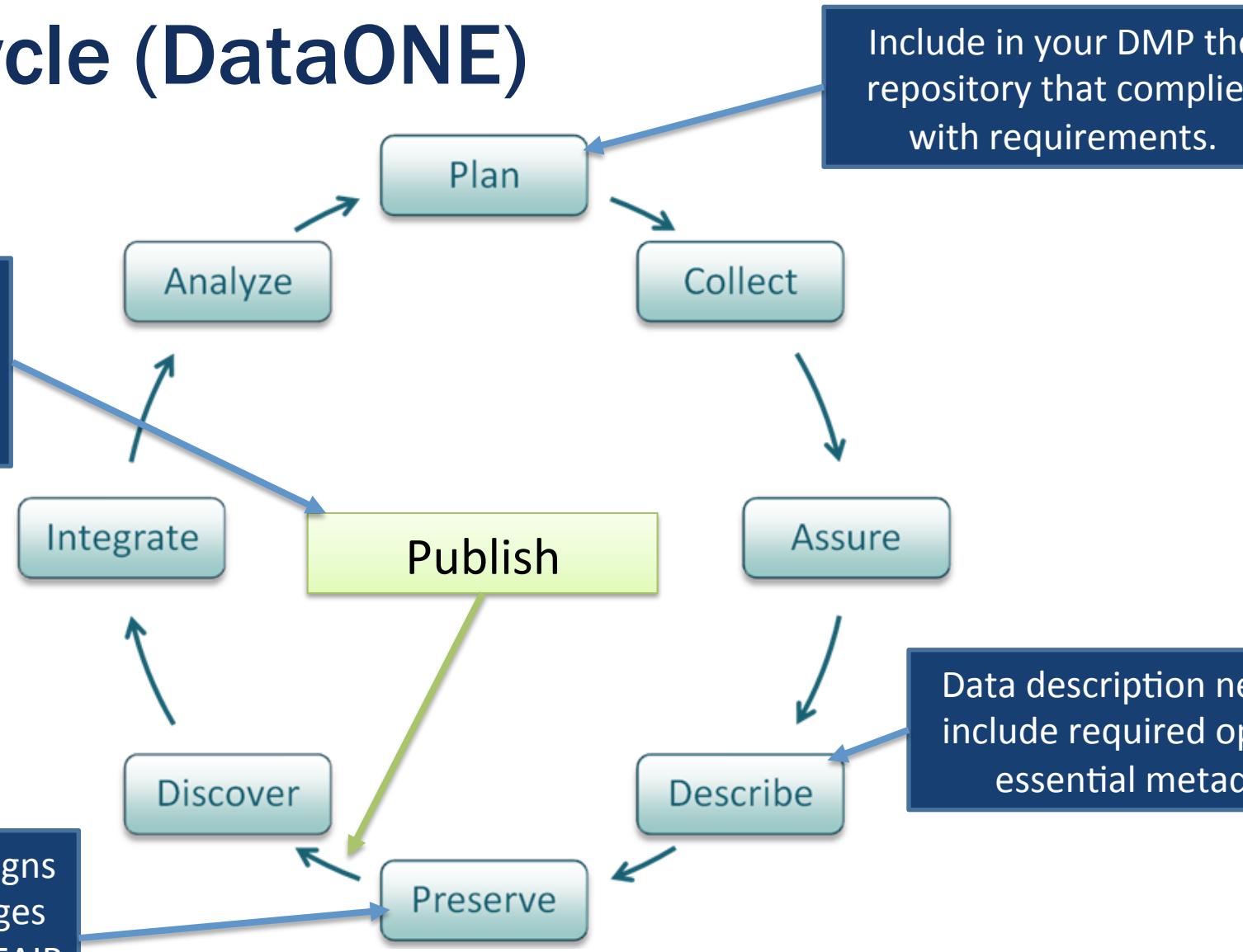
Take Aways...

- 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

Data Lifecycle (DataONE)

Give authors an option to submit data to an appropriate repository at the time of paper submission.

Partner with a repository that assigns DOIs, provides proper landing pages supporting citation, and supports FAIR Guidelines.



Community-Driven Project – Partnership Includes:

- **Science Data Communities**

- AGU
- Earth Science Information Partners (ESIP)
- Research Data Alliance (RDA)

- **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!!

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 – June 2018
Second Stakeholder Meeting	June 2018
- Adoption and Implementation of Guidelines, Recommendations, and Policies	June 2018 – Feb 2019

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 Principles In the Rest of the Lifecycle

- Incentives
- Communication
- Alignment

estions?

Contact Information:

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

MelleyStall
MelleyStall@agu.org

