Enabling FAIR Data:

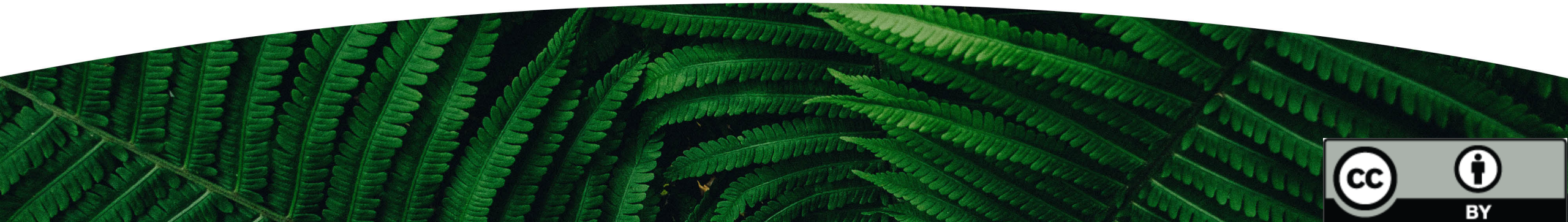
Helping our Researchers Share Their Data and the Reward of Attribution and Credit

19 February 2019

Shelley Stall, AGU Sr. Director, Data Leadership

sstall@agu.org **@ShelleyStall**

https://orcid.org/0000-0003-2926-8353



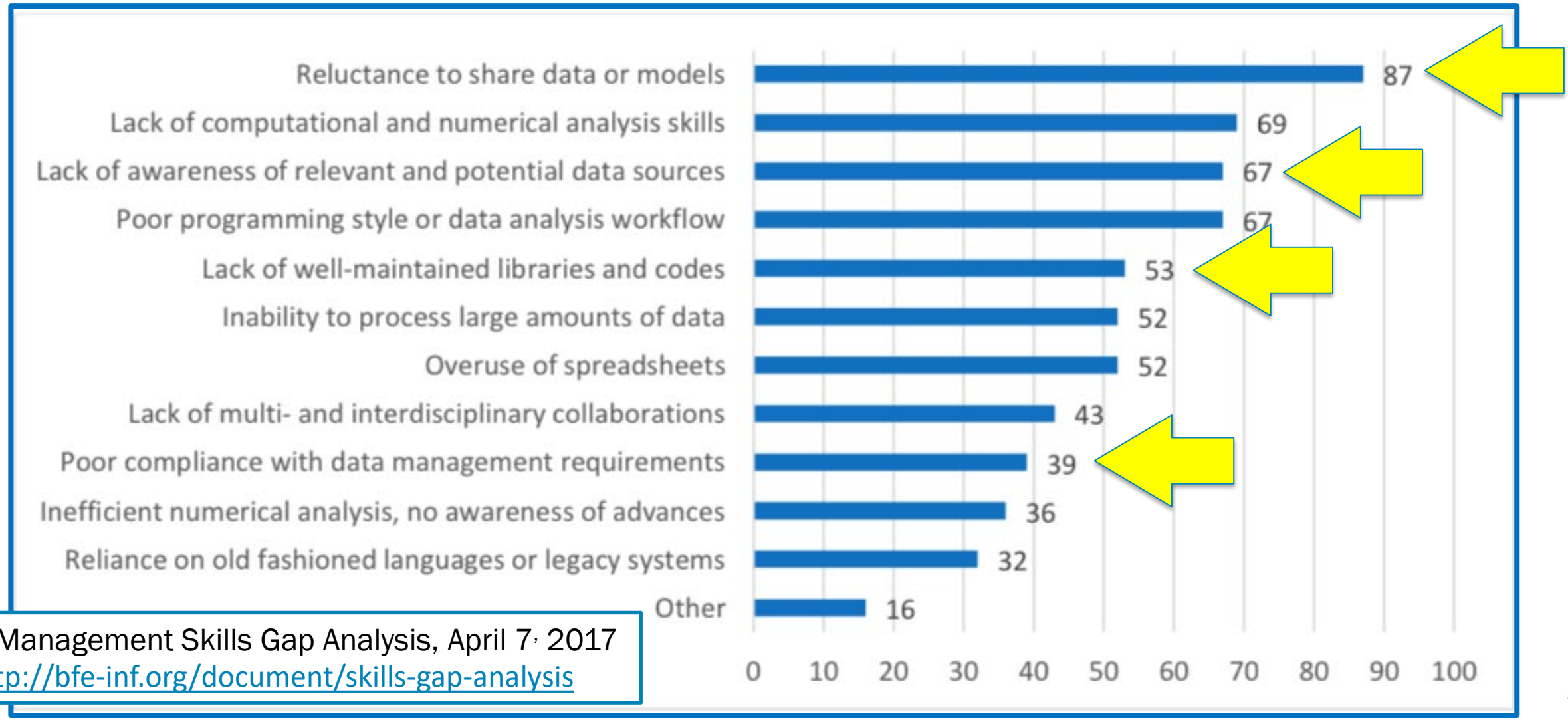
AGU's position statement on data affirms that

"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

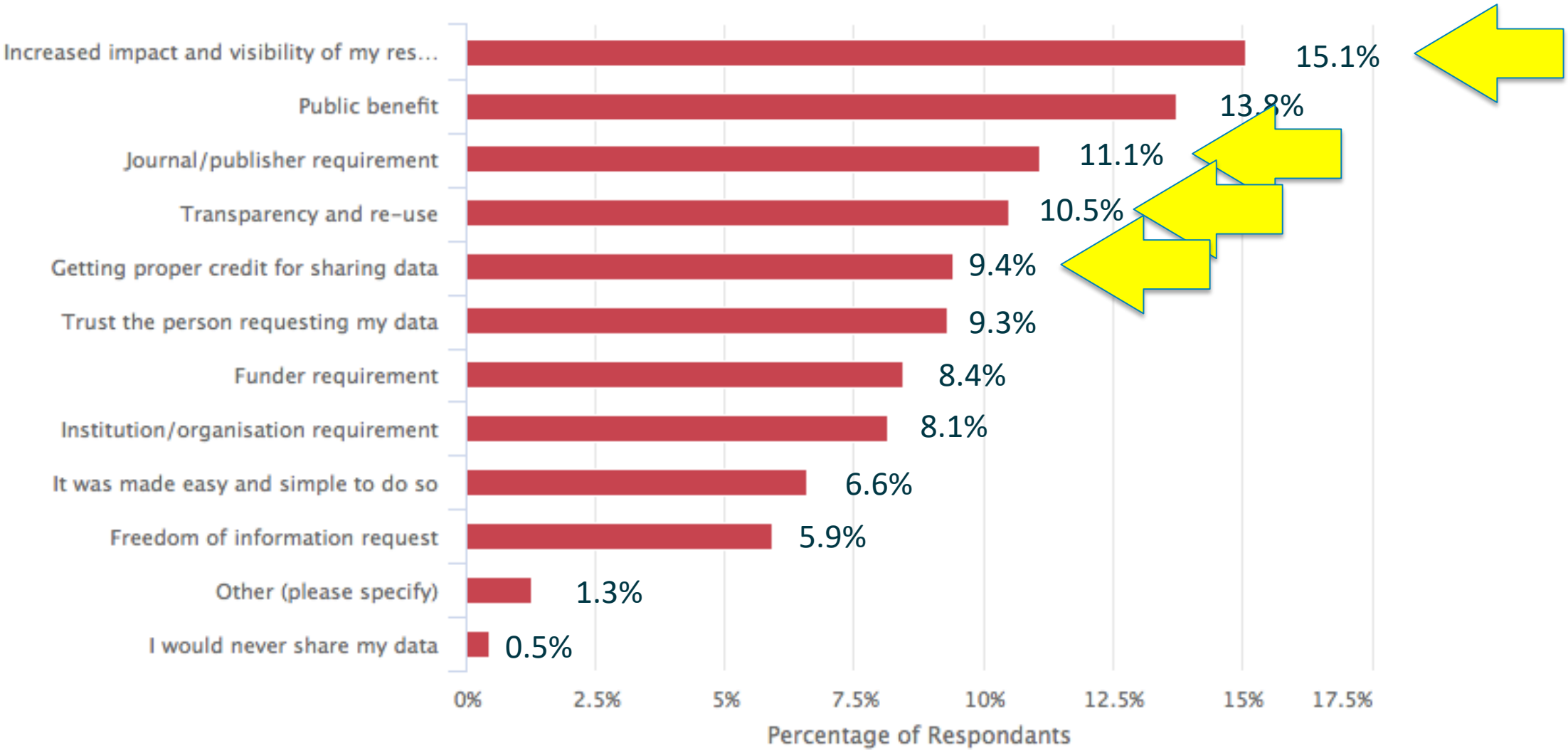
Photo Credit: Photo by Rick Meyers on Unsplash

Research habits and digital skills needing most improvement



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

What circumstances would motivate you to share your data?



https://knowledge.figshare.com/articles/item/state-of-open-data-2017, accessed Dec 9, 2018

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*

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. Rich data documentation (metadata).

Same for any relevant software.

Supported by the repository

Accessible

The data and data documentation (metadata) are accessible.

Supported by the repository

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.

Supported by the repository

Reusable

The data licensing is clear and accessible. Preferably as open as possible.

Supported by the repository


As Strong As The Wind
As Bright As The Sun
And As Deep As The Ocean
Don't Forget to Deposit Your Data With US!!
Love, Your
Repository Friends
https://www.hallmarkecards.com/ecards/collections/thinking+of+you

FAIR Guiding Principles

FAIR is…

Findable

Accessible

Interoperable

Reusable

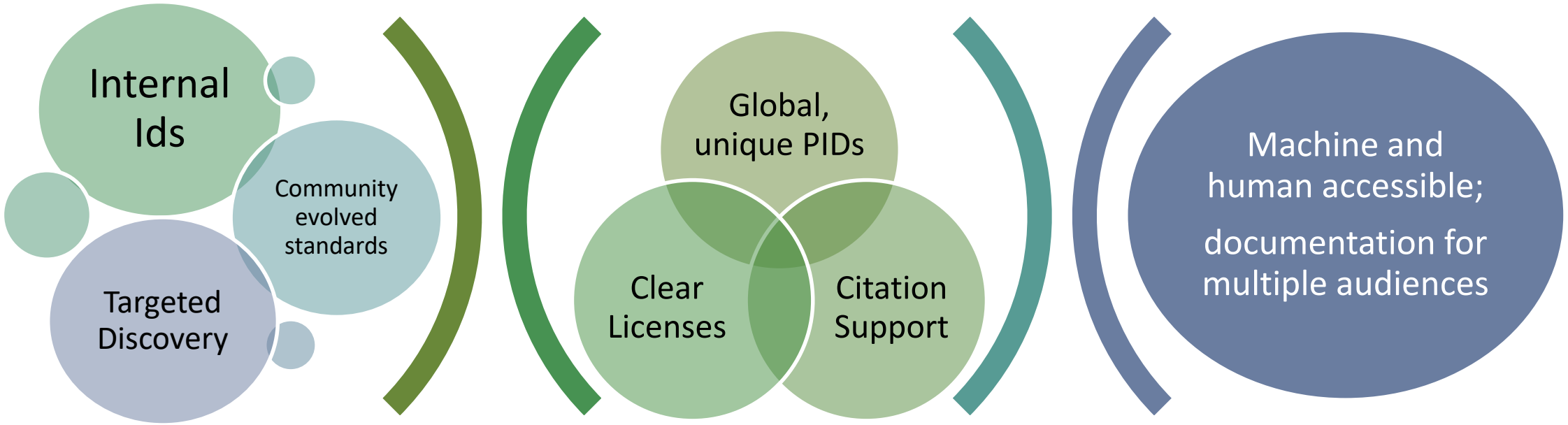
Article in Nature journal *Scientific Data*: Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

FAIR Data Principles (applies to software too)

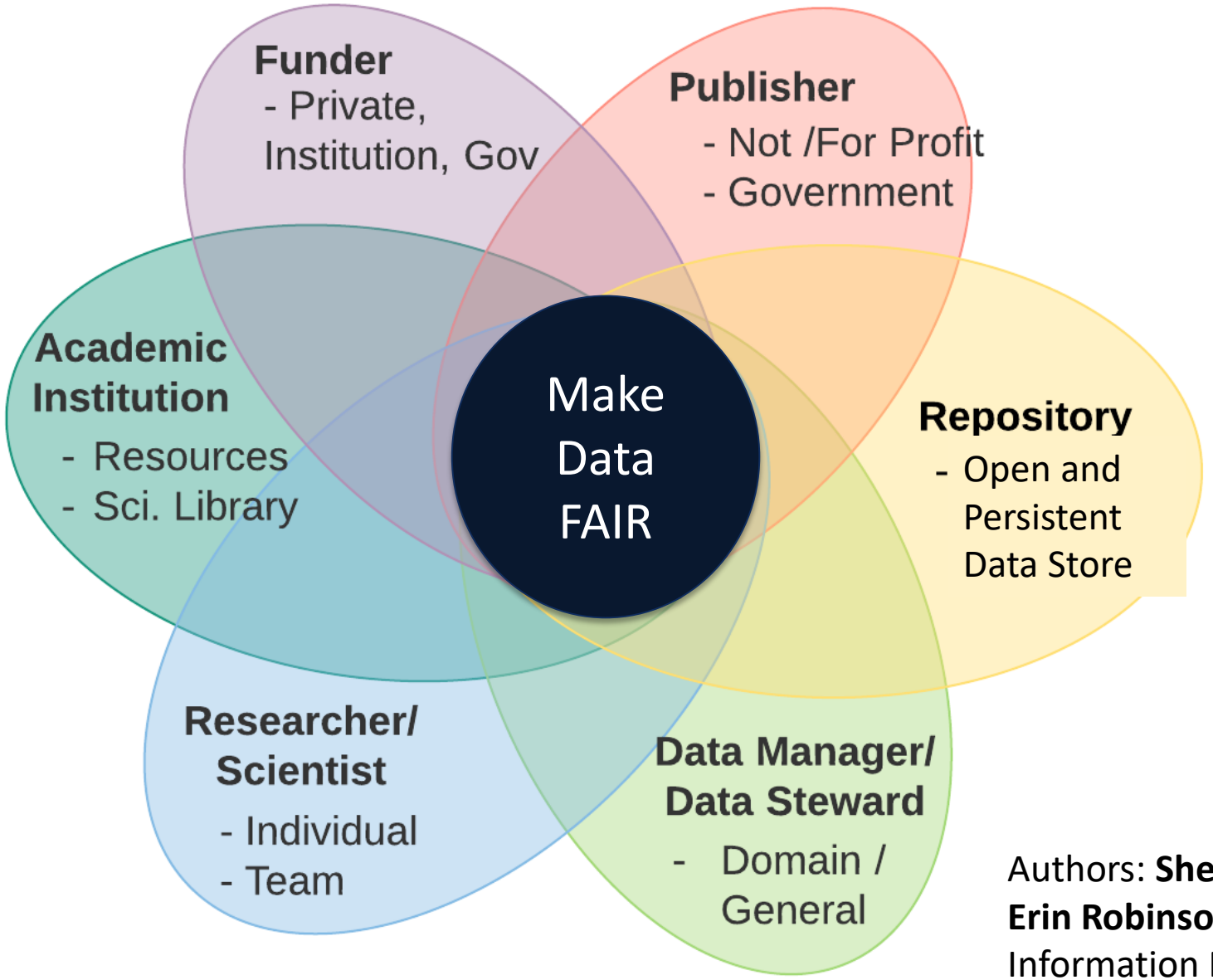
- **Findable**
  - Assign persistent IDs (PIDs), provide rich metadata, register in a searchable resource, …
- **Accessible**
  - Retrievable by their ID using a standard protocol, metadata remain accessible even when data are no longer available…
- **Interoperable**
  - Use formal, broadly applicable 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



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

https://copdess.org/enabling-fair-data-project/

ENABLING FAIR DATA PROJECT

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

READ MORE

Publication of scholarly articles in the Earth, space, and environmental science community is conditional upon the concurrent

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.

Current Publisher Signatories… (as of 18 Feb 2019)

- American Geophysical Union
- Copernicus Publications
- Ubiquity Press
- California Digital Library – CDL
- Wiley
- PANGAEA, Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI), Center for Marine Environmental Sciences, University of Bremen (MARUM)
- WDC Climate, Deutsches Klimarechenzentrum (DKRZ)
- *Science*
- *Science Advances*
- PLOS
- Elsevier
- F1000
- *Nature*
- *Scientific Data*

Over 100 signatories to date!

https://copdess.org/enabling-fair-data-project/commitment-statement-in-the-earth-space-and-environmental-sciences/signatories/

FAIR-Aligned: Researcher Commitment

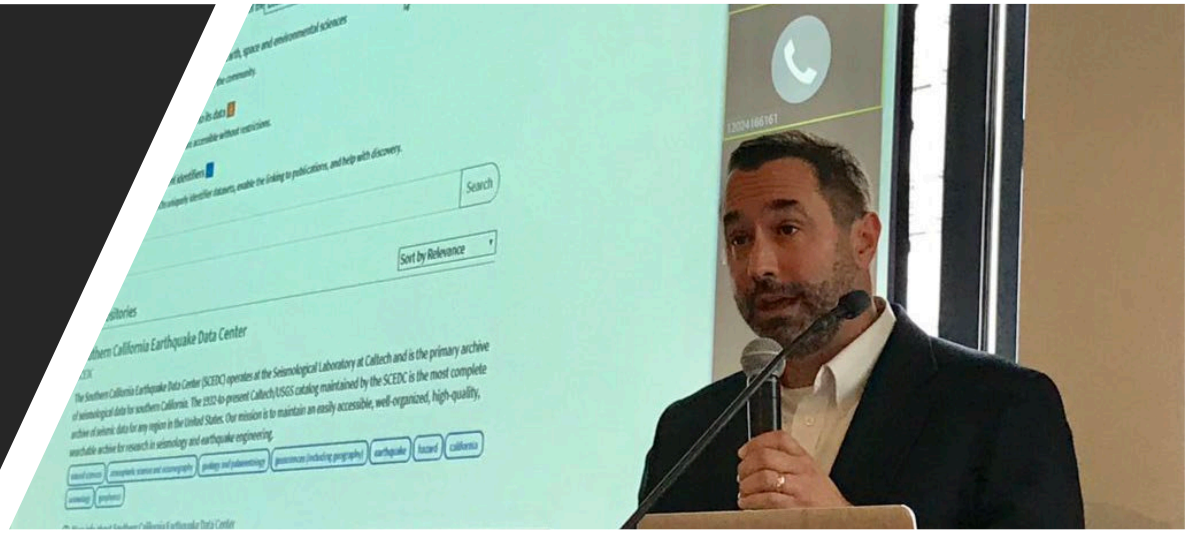
- Locating trustworthy, community-accepted, FAIR-aligned repositories that support:
  - Documenting data and software (and other research outputs as is possible) to agreed community standards that describe provenance and enable discovery, assessment of reliability, and reuse
  - Persistent identifiers 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.

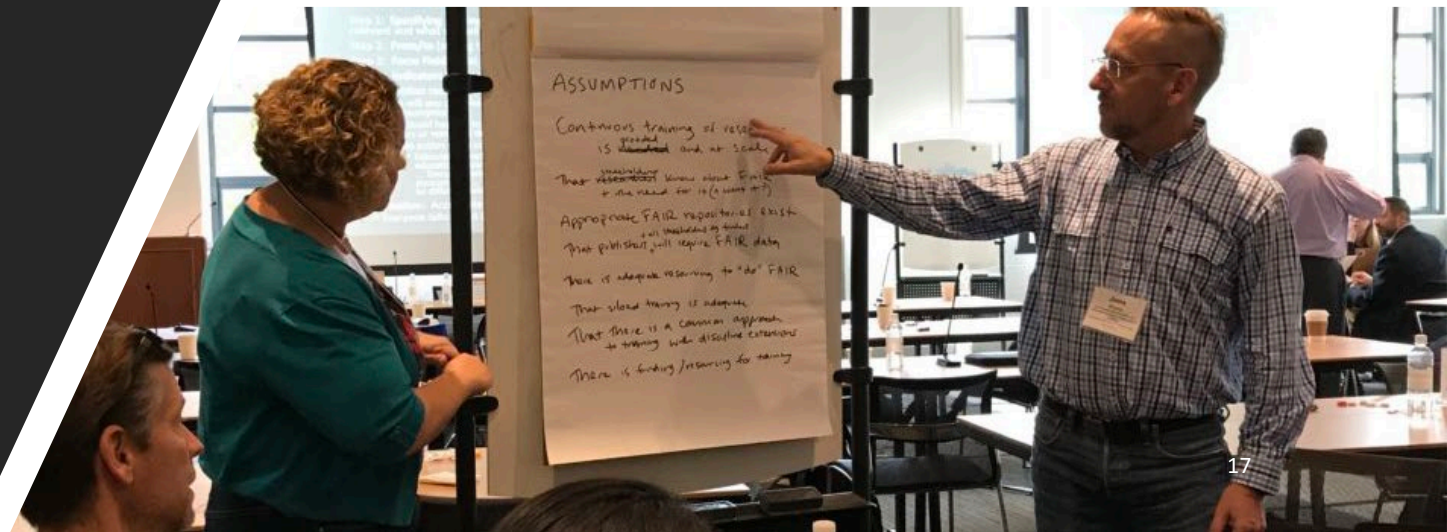
FAIR-aligned: Repository Commitment

- 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.

Data Sharing and Citations: New Author Guidelines

Stall, S., et al. (2018), Data sharing and citations: New author guidelines promoting open and FAIR data in the Earth, space, and environmental sciences, *Sci. Editor,* https://www.csescienceeditor.org/article/data-sharing-and-citations-new-author-guidelines-promoting-open-and-fair-data-in-the-earth-space-and-environmental-sciences/.





FAIR-Aligned Publishers - Author Guidelines for Data (1 of 2)

- **Deposit research data in a FAIR-aligned repository,** with a preference for those that explicitly follow the FAIR Data Principles and demonstrate compliance with international standards for data repositories, (**e.g.** CoreTrustSeal).
- **Cite and link to the data** in the article, following the Joint Declaration of Data Citation Principles and ESIP Guidelines, using the unique, resolvable, and persistent identifiers provided by the repository in which the data are archived.
- **Include a Data Availability Statement** describing how the data underlying the findings of their article can be accessed and reused.

FAIR-Aligned Publishers - Author Guidelines for Data (2 of 2)

- **Provide unrestricted access to all data** and materials underlying reported findings for which ethical or legal constraints do not apply.
- A **tool to assist in identifying FAIR-aligned repositories** is available from DataCite and can be found at https://repositoryfinder.datacite.org .
- There may **be a need to restrict some access to data** because of fragile environments, endangered species, geopolitical tensions or cultural sensitivities (e.g. indigenous land rights).

Developed by: DataCite/re3data https://repositoryfinder.datacite.org

Repository Finder

Find a repository to upload your data.

Repository Finder, a pilot project of the Enabling FAIR Data Project led by the American Geophysical Union (AGU) in partnership with DataCite and the Earth, space and environment sciences community, can help you find an appropriate repository to deposit your research data. The tool is hosted by DataCite and queries the re3data registry of research data repositories.

Search re3data for a repository to upload your data

Type to search... Search

or

See the repositories in re3data that meet the criteria of the Enabling FAIR Data Project.

Repository Finder

About Search FAQ Support Services

Search re3data for a repository to upload your data

See the repositories in re3data that meet the criteria of the Enabling FAIR Data Project

Type to search... Search

187 Repositories

Sort by Name

4TU.Centre for Research Data

4TU.ResearchData

4TU.ResearchData, previously known as 3TU.Datacentrum, is an archive for research data. It offers the knowledge, experience and the tools to share and safely store scientific research data in a standardized, secure and well-documented manner. 4TU.Centre for Research Data provides the research community with: Advice and support on data management; A long-term archive for scientific research data; Support for current research projects; Tools for reusing research data.

life sciences | basic biological and medical research | biophysics | bioinformatics and theoretical biology | agriculture, forestry, horticulture and veterinary medicine | soil sciences | biology | agriculture, forestry, horticulture and veterinary medicine | natural sciences

Feedback

Repository Finder

About Search FAQ Support Services

Search re3data for a repository to upload your data

See the repositories in re3data that meet the criteria of the Enabling FAIR Data Project

The repository is a domain repository for the Earth, space and environmental sciences

The repository has special expertise and is used by the community.

The repository provides open access to its data

Research data hosted by the repository are accessible without restrictions.

The repository uses persistent identifiers

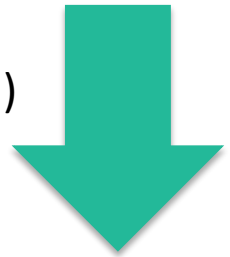
Persistent identifiers such as DOIs uniquely identifier datasets, enable the linking to publications, and help with discovery.

Change criteria for this search

Type to search...

Search

Indication of Certification (e.g. CoreTrustSeal)



National Geoscience Data Centre

NGDC

The BGS is a data-rich organisation with over 400 datasets in its care; including environmental monitoring data, digital databases, physical collections (borehole core, rocks, minerals and fossils), records and archives. Our data is managed by the National Geoscience Data Centre.

natural sciences | atmospheric science and oceanography | geophysics and geodesy | geophysics | water research | geosciences (including geography) | hydrology | groundwater levels | climate change | earthquakes | geothermal energy | materials collections | offshore oil | offshore gas | drilling | borehole records

More info about National Geoscience Data Centre

Frequently Asked Questions

https://copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/

Table of Contents

ESES CoreTrustSeal Cohort - Formed

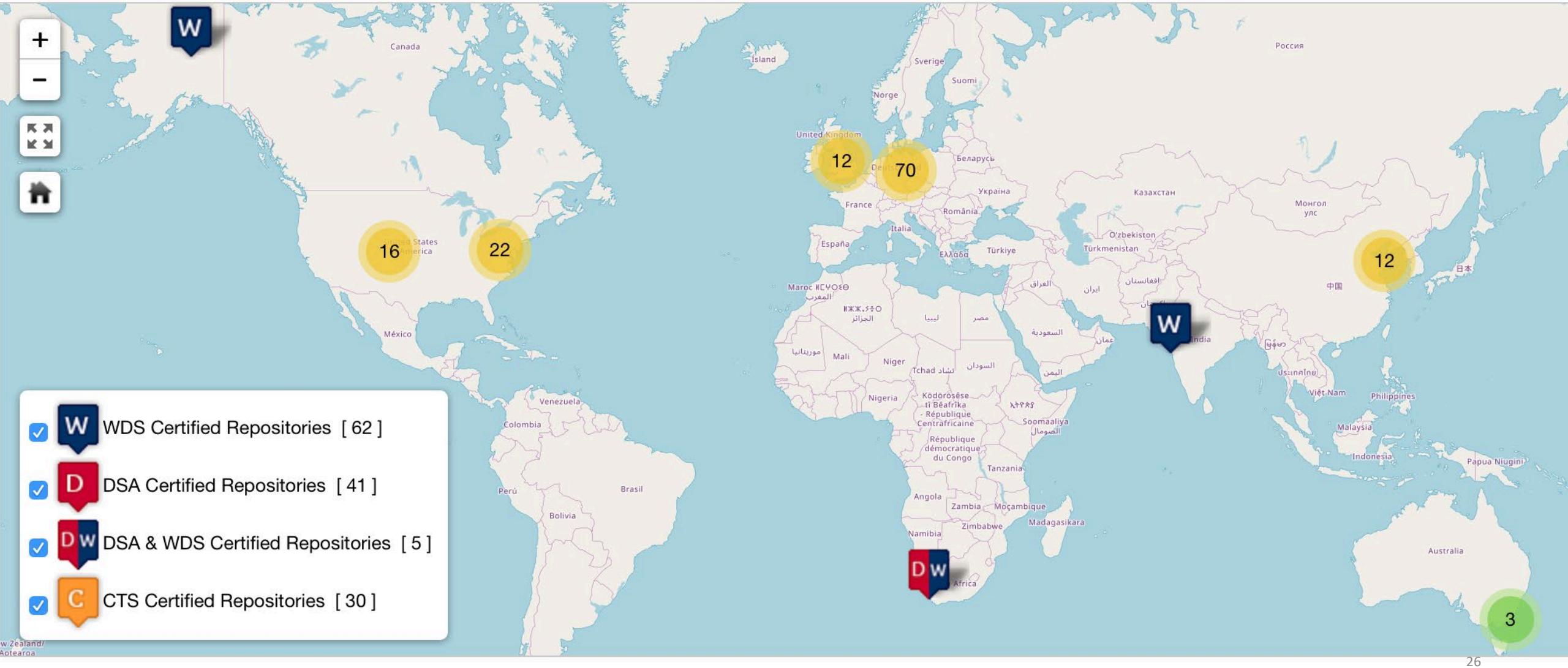
- No Risk to Join the Cohort
- Repositories specific to the Earth, space, and environmental science community
- Webinars recorded to you can catch up easily – YouTube.com, EarthCubeNSF Channel
  - https://youtu.be/gs9HTzxbx0s, 30 October 2018
    - Rorie Edmunds, World Data System Programme Officer
    - John Faundeen, USGS EROS Archivist
  - https://youtu.be/vV_3ZcDsagA, 26 November 2018
    - Ingrid Dillo, Deputy Director of DANS, CoreTrustSeal Board Member
    - Helen Glaves, Senior Data Scientist at the British Geological Survey, UK National Geoscience Data Centre
- **Initial cost of CoreTrustSeal certification application is covered for the first 25 repositories that submit packages.**

If you are interested in becoming a member of the cohort, please send email to Rebecca Koskela (rkoskela@unm.edu).

Also, please pass this on to anyone you think would be interested in joining the cohort.

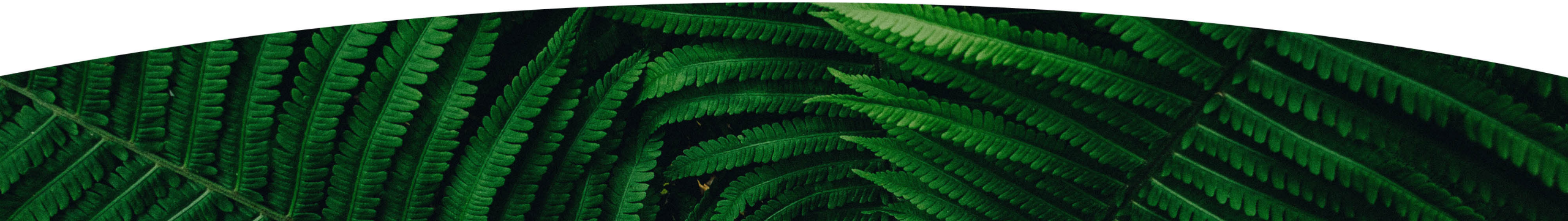
CoreTrustSeal.org Certification

ICSU World Data System (WDS)



Author, Reviewer, Editor – Open and FAIR – New Data Guidelines – One-Page Reference

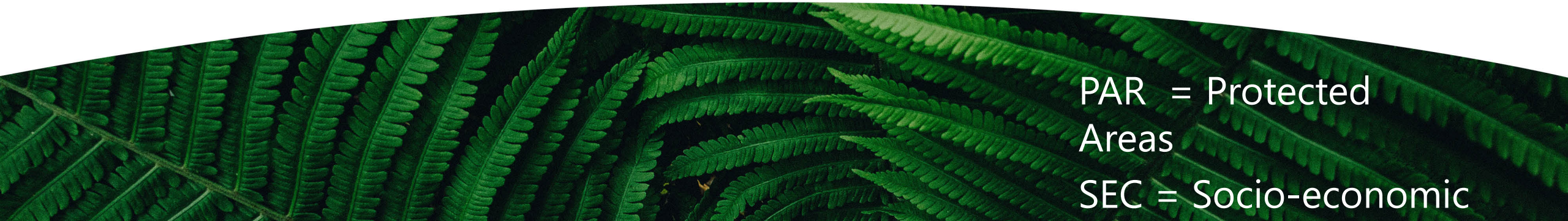
1. Summary of objectives for open and FAIR data
2. Resources for Editors, Reviewers, and Authors
3. Information and description of the tools for editors, reviewers, and researchers
4. Data Citation Training, Guidelines, and Examples
5. Software Citation Guidelines and Examples
6. Enabling FAIR Data Project - Overview
7. FAIR Guiding Principles



Belmont Forum

Collaborative Research Action (CRA), Science-driven e-Infrastructure Innovation (SEI) for the Enhancement of Transnational, Interdisciplinary and Transdisciplinary Data Use in Environmental Change

PARSEC - Building New Tools for Data Sharing and Re-use through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas

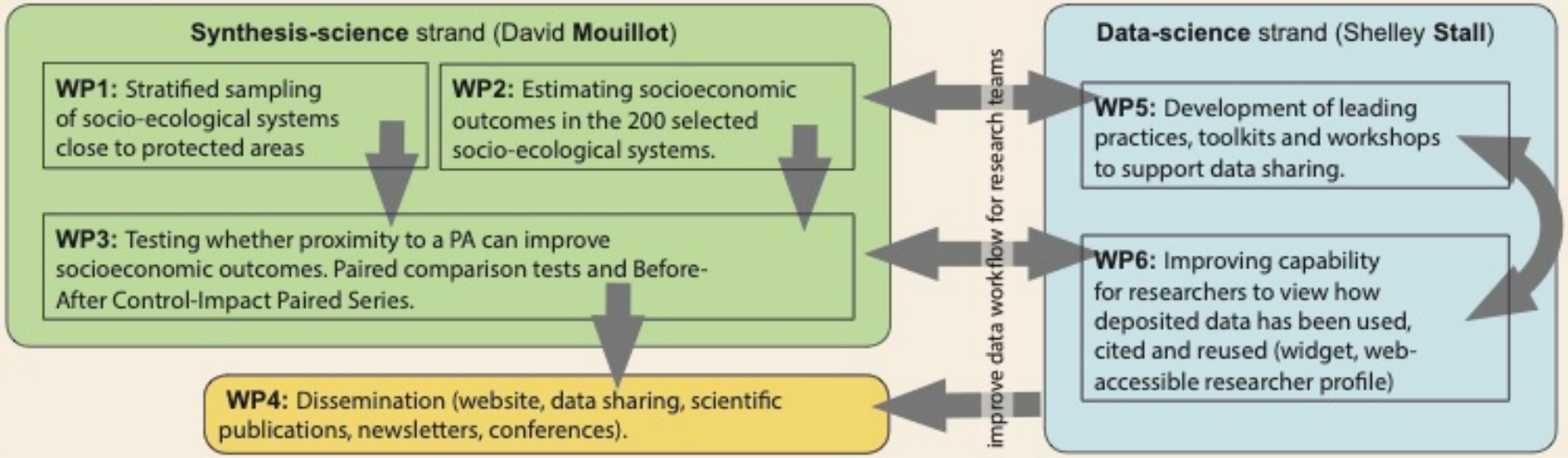


PARSEC : Building New Tools for Data Sharing and Reuse through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas

Consortium Leaders: Nicolas **Mouquet** , David **Mouillot**, Alison **Specht** and Shelley **Stall**.

Objectives

(a) Determine the influence of natural protected areas (PAs) on the socio-economy of local communities;
(b) Develop metrics for prediction and mitigation of adverse effects of PAs;
(c) Improve linkage between data, publications and researchers;
(d) Improve recommendations for the research data workflow and skills for research teams;
(e) Improve future environmental decision-making;
(f) Increase the number of citations to data sets and better attribute them to the data creator;
(g) Promote the incentive of credit for open and FAIR data management and preservation for data reuse;
(h) Provide tools for researchers to view how the data they have deposited is cited and reused.



FUNDING: 1258K€

DURATION: 48 months

FRANCE: Foundation for Research on Biodiversity - ANR 500 K€ (N. **Mouquet**)
USA: American Geophysical Union - NSF 425 K€ (S. **Stall**)
JAPAN: National Institute of Information & Communications Technology - JST 249 K€ (Y. **Murayama**)
BRAZIL: Universidade de São Paulo, Brazil - FAPESP 84 K€ (P. Pizzigatti **Corrêa**) plus: XXX
Collaborative partners: NCI, Australia (L. Wyborn), BGU, UK (H. Glaves). **Associates:** DataCITE, ORCID, ESIP, RDA, LTER

USA Team Members

- **Shelley Stall - Country Team lead and Lead PI for data-science.**
  - **Synthesis-science**
    - Jamie Trammell (country team leader): natural resource planning for conservation.
    - Robin Chazdon: biodiversity and human well-being in tropical forested areas.
    - David Kramer: on-ground community engagement in developing countries.
    - To be determined - technician/intern
  - **Data-science**
    - Shelley Stall, (country team leader and Lead PI for Data-science): data sharing, reuse, management, attribution, and process improvement. FAIR data expert. Scholarly publishing and scientific society.
    - Margaret O'Brien: semantics and management of heterogeneous data.

Questions?

Shelley Stall

AGU Senior Director, Data Leadership

sstall@agu.org @ShelleyStall

https://orcid.org/0000-0003-2926-8353