FAIR-aligned Scientific Repositories: Essential Infrastructure for Open and FAIR Data

GeoDaRRs: What is the existing landscape and what gaps exist in that landscape for data producers and users?

7 August 2018

Shelley Stall, AGU Director, Data Programs

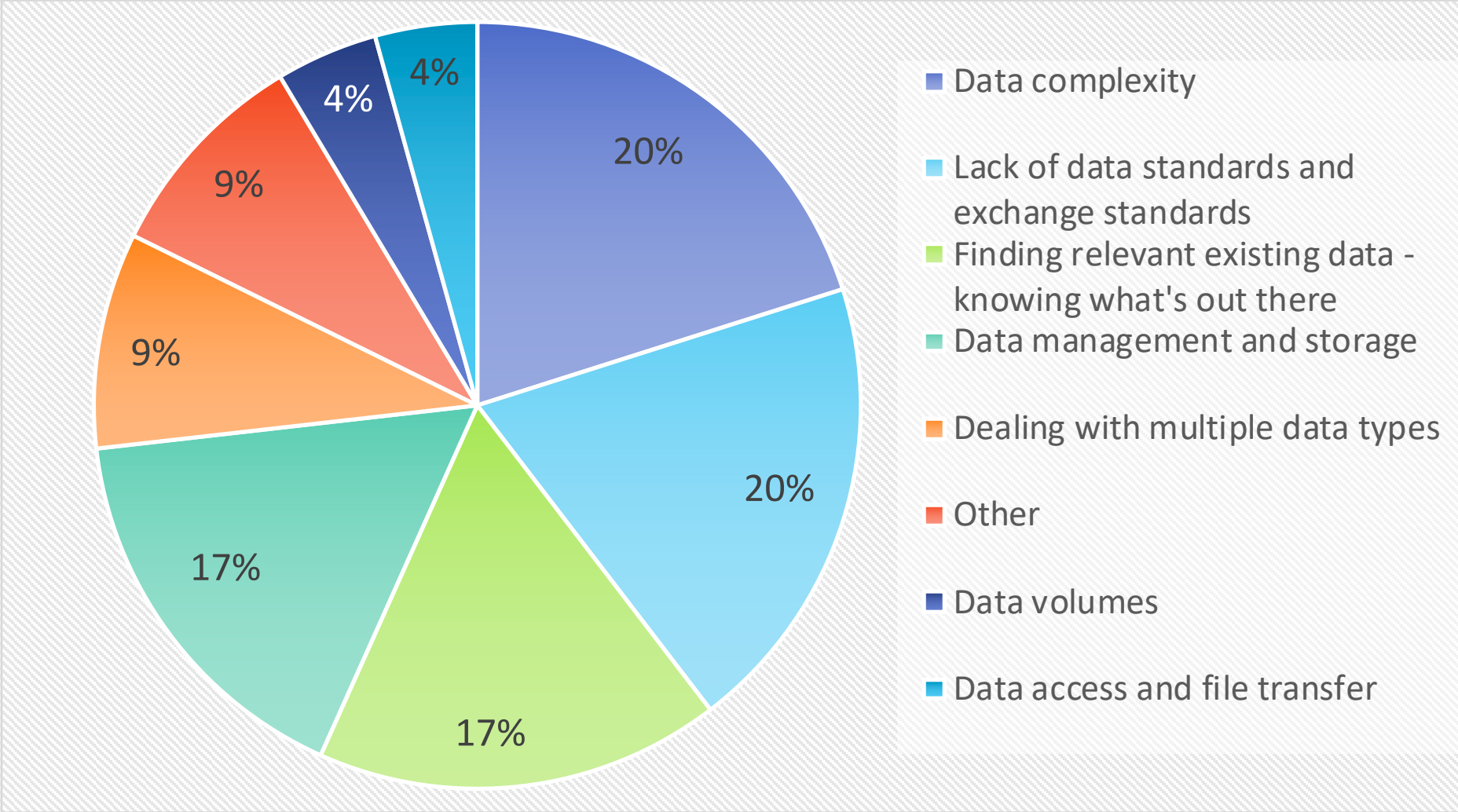**sstall@agu.org   @ShelleyStall**

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."

Researcher Challenges with Data Use

The top four issues accounted for 73% of respondents

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

Pie chart values: 20%, 20%, 17%, 17%, 9%, 9%, 4%, 4%

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 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
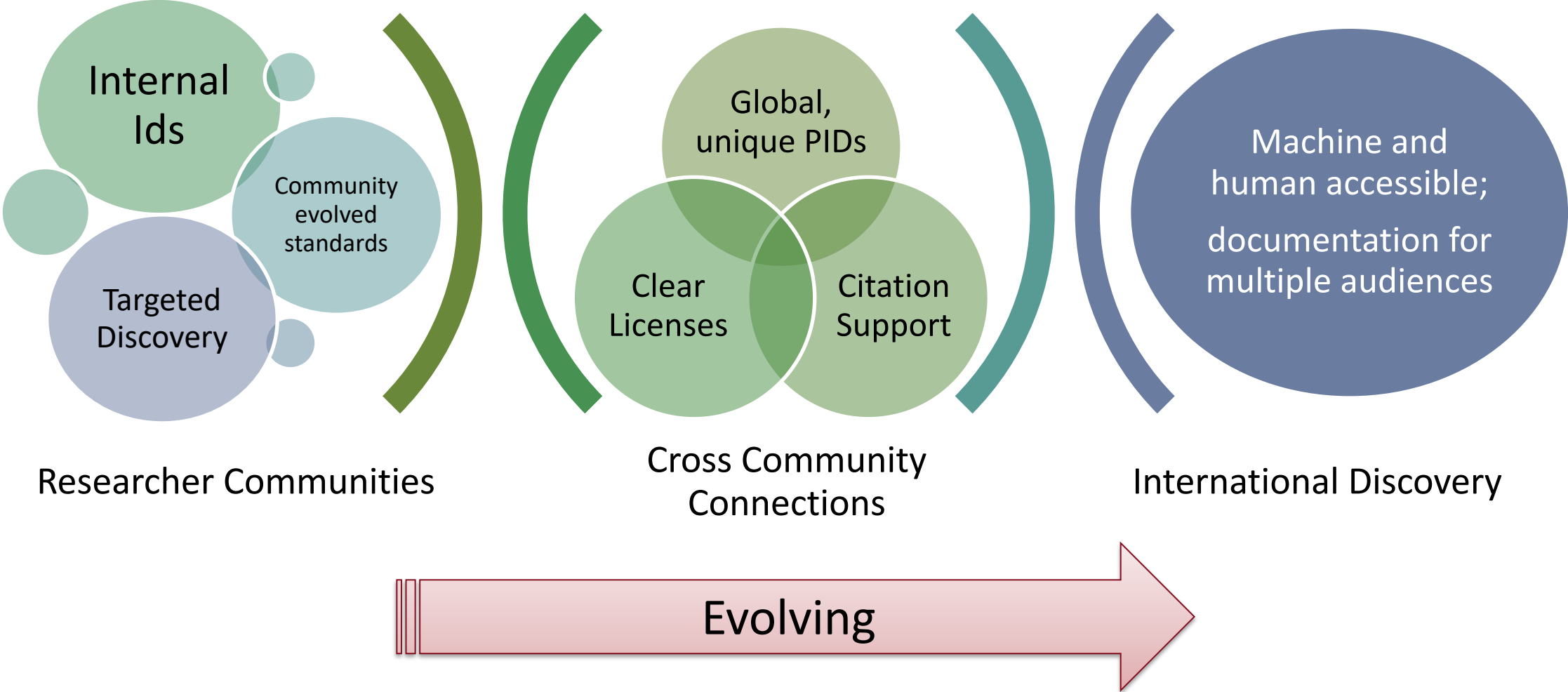 - Use formal, broadly applicable languages, use standard vocabularies, qualified references...

- ## 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.

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

COPDESS
Coalition for Publishing Data in the Earth and Space Sciences

The Coalition for Publishing Data in the Earth and Space Sciences ⌄ | **Enabling FAIR Data Project** ⌄ | 🔍

ENABLING FAIR DATA PROJECT

HOME / ENABLING FAIR DATA PROJECT

Funded by the Laura and John Arnold Foundation

OVERVIEW | LEADERSHIP | COMMITMENT STATEMENT | RESOURCES | FAQS

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.**

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.

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!!

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

Thank you!

Shelley Stall

Director, AGU Data Program

sstall@agu.org

@ShelleyStall