TECH NOTES

Facilitating Connections: The CI Compass FAIR Data Topical Working Group

Author(s): Angela P. Murillo, Don Brower, and Charles F. Vardeman II

The U.S. National Science Foundation (NSF) CI Compass Cyberinfrastructure Center of Excellence's mission is to foster discussion and provide expertise and support to cyberinfrastructure practitioners at NSF Major and Mid-scale Facilities. One issue that is important across the facilities is FAIR (Findable, Accessible, Interoperable, Reusable) Data. In response to feedback from the 2022 Cyberinfrastructure for NSF Major Facilities (CI4MF) Workshop in which the Major Facilities (MFs) and broader cyberinfrastructure (CI) community shared that they would like more guidance regarding FAIR Data, CI Compass created its FAIR Data Topical Working Group (TWG) in August 2022.

The FAIR Data TWG aims to:
- understand the current practices and needs of FAIR implementation by MFs,
- research current and emerging FAIR data practices and implementations,
- organize guest speakers who have experience and expertise with FAIR, and
- disseminate research regarding FAIR data practices and implementation.

The NSF CI Compass FAIR Data TWG includes CI Compass personnel, MF personnel, and colleagues from the broader CI Community and is open to all MF personnel and CI colleagues interested in participating.

Year One FAIR Data TWG Activities

During Year 1, between 2022 and 2023, the CI Compass FAIR Data TWG has:
- Solicited topics and organized monthly meetings based on those topics,
- Researched and documented considerations for implementing FAIR,
- Organized guest speakers for FAIR-related webinars, and
- Disseminated research regarding FAIR implementation.

In January 2022, a summary of the FAIR principles, how FAIR is related to the data lifecycle, and how FAIR impacts major facilities was described in the CI Compass Tech Note "Making the Major Facilities Data Lifecycle FAIR."

Additionally, in January 2023, the FAIR Data TWG organized a webinar that brought together Martin Halbert, the Advisor for Public Access at the NSF, and Shawna Sadler the Head of Outreach and Partnerships at ORCID to discuss the 2022 White House Office of Science and Technology Policy (OSTP) memo, known as the "Nelson Memo," persistent identifiers, and how they impact research facilities. The webinar is publicly available through the CI Compass YouTube page here.

Joint Conference on Digital Libraries (JCDL) Workshop

To create a space for further discussion regarding FAIR Data and MFs, the FAIR Data TWG organized a workshop at the JCDL Workshop hosted by the Los Alamos National Laboratory in June 2023.

The 2023 JDCL Workshop "FAIR Data for Large Research Facilities," brought together data managers, repository managers, administrators, and others who are responsible for or interested in research data management at large research facilities.

This workshop aimed to provide cross-pollination between facilities that have similar desires to realize the FAIR principles and included discussion on issues related to data management that are of interest to large research facilities.

Discussions covered many topics in FAIR and cybersecurity, including the Nelson Memo and open science support. The group heard about a desire to support users effectively on cloud platforms, especially in terms of making data convenient to use. This necessitates using third-party platforms, such as Google Cloud Platform or AWS. Most interesting was a novel integration with the Open Science Framework to smooth the transport of data from the facility to researchers' labs for analysis. Facilities were thinking about the community using their open data and how to support them better and grow the community. Open questions included ways to provide for data storage post-award, and how to do this efficiently within existing funding, which is challenging with the growing volume of data.

FAIR Data TWG FAIR Data Survey

During the Spring of 2023, the FAIR Data TWG created and disseminated a survey to understand the current status of FAIR data implementation at MFs.

The 15-question survey included questions regarding:
- Familiarity with FAIR
- Progress toward FAIR implementation
- Barriers and non-barriers to FAIR implementation
- The value of FAIR implementation
- What is needed to improve FAIR implementation

Participants were recruited through the NSF - Research Infrastructure Office (RIO) Communities of Interest portal and CI Compass listservs.

Preliminary Results

The survey received 44 complete responses. Participants included personnel from various MFs and other large-scale scientific data facilities. Survey responders included managers/supervisors (48.1%), researchers (14.8%), data specialists (11.1%), repository Managers (5.6%), and "Other" (i.e., data curators, data scientists, and data managers) (9.3%) from 15 major facilities.

These MFs included:
1. U.S. Antarctic Program
2. Arecibo Observatory
3. Academic Research Fleet
4. IceCube Neutrino Observatory
5. International Ocean Discovery Program
6. Large Hadron Collider
7. Laser Interferometer Gravitational-wave Observatory
8. National Ecological Observatory Network
9. National High Magnetic Field Laboratory
10. National Optical-Infrared Astronomy Research Laboratory
11. National Radio Astronomy Observatory

12. [Seismological Facility for the Advancement of Geoscience/Geodetic Facility for the Advancement of Geoscience](#)
13. [Ocean Observatories Initiative](#)
14. [National Hazards Engineering Research Infrastructure](#)
15. [Cornell High Energy Synchrotron Source](#)

Additionally, several mid-scale facilities responded, and research computing centers participated as well.

Most participants considered themselves either extremely or very familiar with FAIR, while only 2.4% considered themselves not familiar with FAIR at all.

With regard to communicating the importance of FAIR, most facilities have engaged in moderate to significant efforts, with only a small percentage not addressing FAIR. Most have offered broad training; however, in-depth or focused training has been rare. Many facilities have made substantial progress in developing facility-wide plans for implementation. A moderate amount of them created schedules for implementation (See Figure 1).
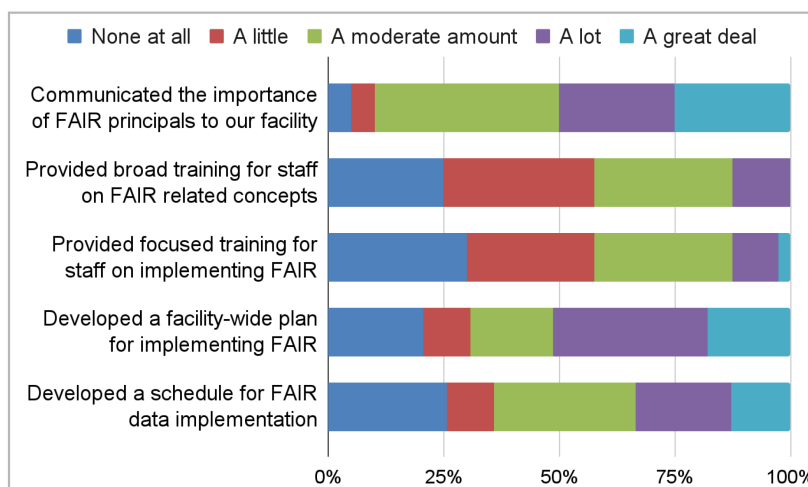

Figure 1. Progress toward FAIR implementation

Most respondents experienced mild to moderate hindrance regarding the lack of tools, while a smaller portion reported little to no hindrance. Most respondents faced minor to moderate challenges with regard to people who understand FAIR. Regarding lack of time prioritizing FAIR data, many participants experienced significant to major obstacles. Regarding the lack of money, respondents reported a wide range of hindrances, from minimal to significant, with a smaller group facing major challenges (Figure 2).
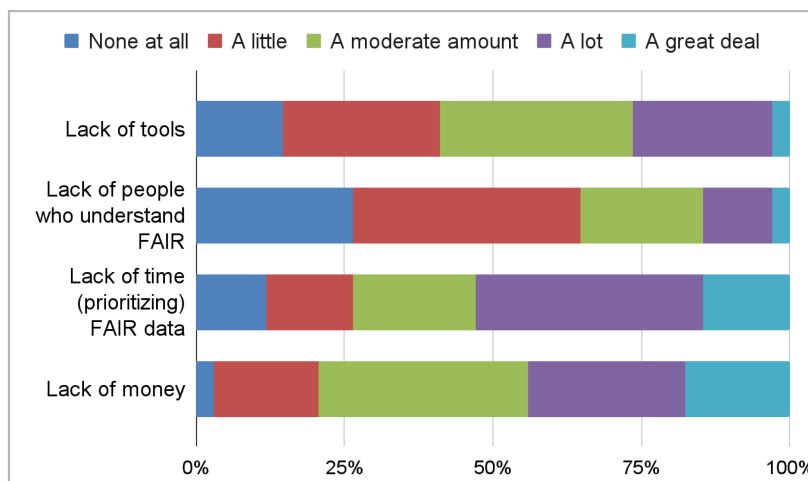

Figure 2. Hindrances for FAIR Implementation

In terms of helpful factors for FAIR implementation for an organization, technical tools are moderate to very beneficial for most participants, with a small portion reporting them as extremely helpful. People who understand FAIR have been predominantly beneficial. Leadership that supports FAIR has shown a significant impact, with a large percentage of respondents finding it very helpful, followed by extremely and moderately helpful. Funding has been moderately helpful for many participants while proving very helpful for others (See Figure 3).
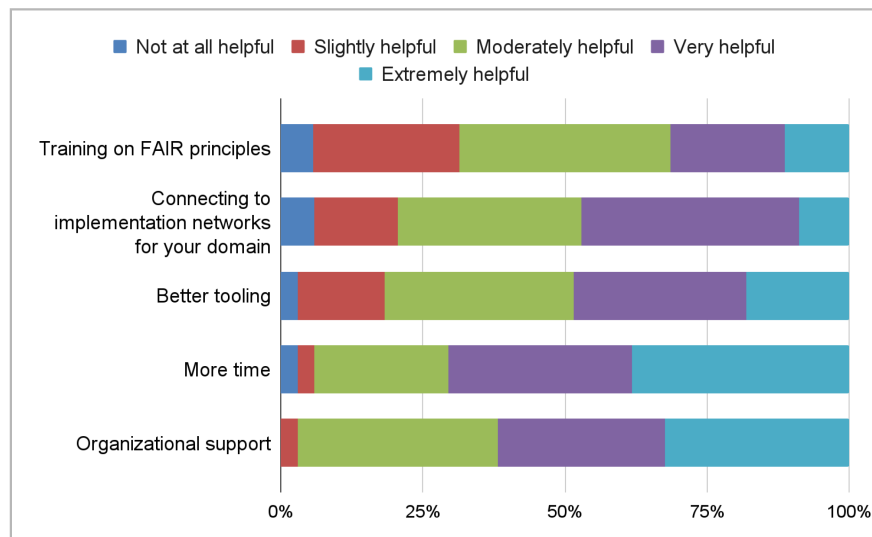

Figure 3. Helpful factors for FAIR implementation (organizationally)

With regard to helpful factors for FAIR implementation for individuals, a majority of respondents found training on FAIR principles, connecting to implementation networks, and better tooling to be moderate to extremely helpful. More time and organizational support were considered very or extremely helpful by most respondents (See Figure 4).
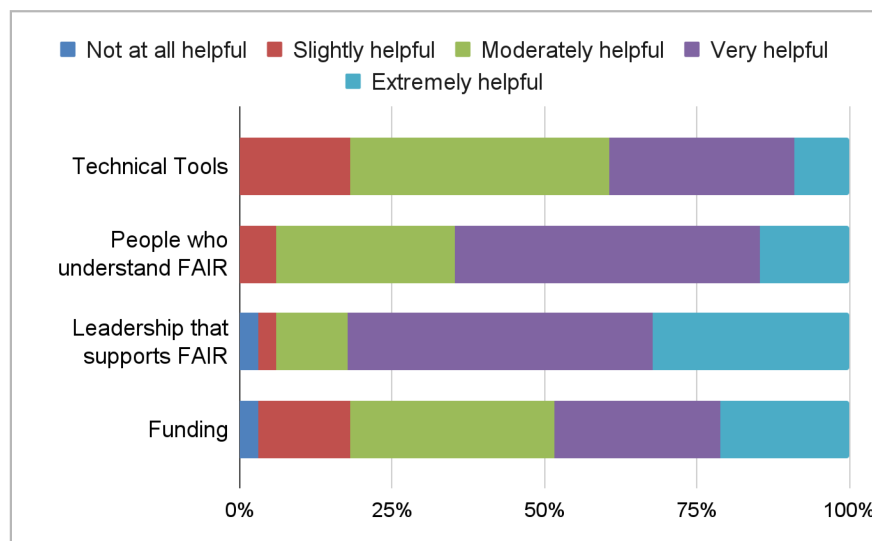

Figure 4.   Helpful factors for FAIR implementation (individually)

Most participants see the value of FAIR both from the perspective of grant requirements and personal value beyond grant requirements. Participants strongly agreed that FAIR practices help their facility, with some also somewhat agreeing.

A substantial number of respondents strongly agreed that FAIR provides value to their discipline, with a smaller group somewhat agreeing. Finally, an overwhelming majority strongly agreed that FAIR provides value to science, highlighting the respondents' overall positive view of FAIR (See Figure 5).
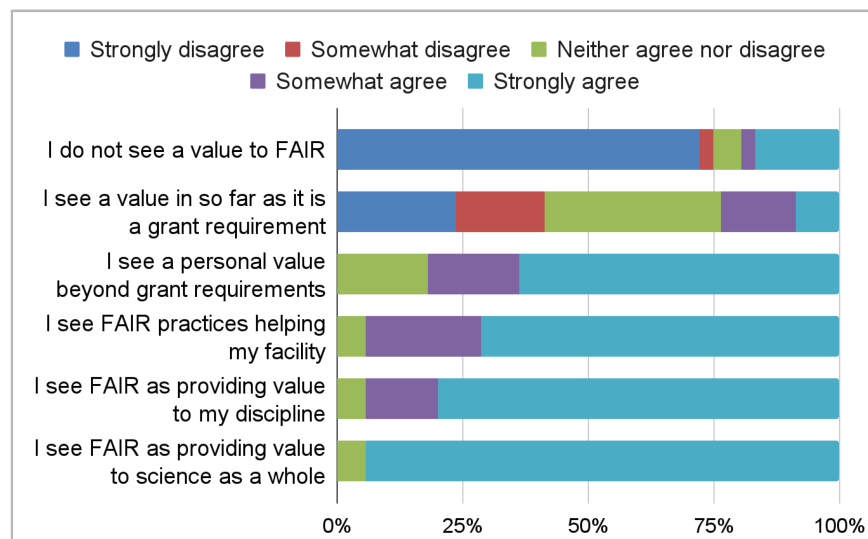

Figure 5. Perceptions of FAIR

The full results of the survey were presented at the [2024 International Digital Curation Conference](#) in February 2024, and the paper proceedings will be published in the [International Journal on Digital Curation](#).

Additionally, the FAIR Data Working Group conducted surveys with nine MF personnel to gain a more nuanced understanding of FAIR data implementation at MFs. These interviews are currently being analyzed, and the results will be disseminated in 2024.

___

If you are interested in the CI Compass FAIR Data Topical Working Group, or would like to learn more, please reach us at contact@ci-compass.org.