

FAIR TWG Next Steps

FAIR Facilities and Instruments — Workshop #2

August 22, 2024

Don Brower, CI Compass / Notre Dame

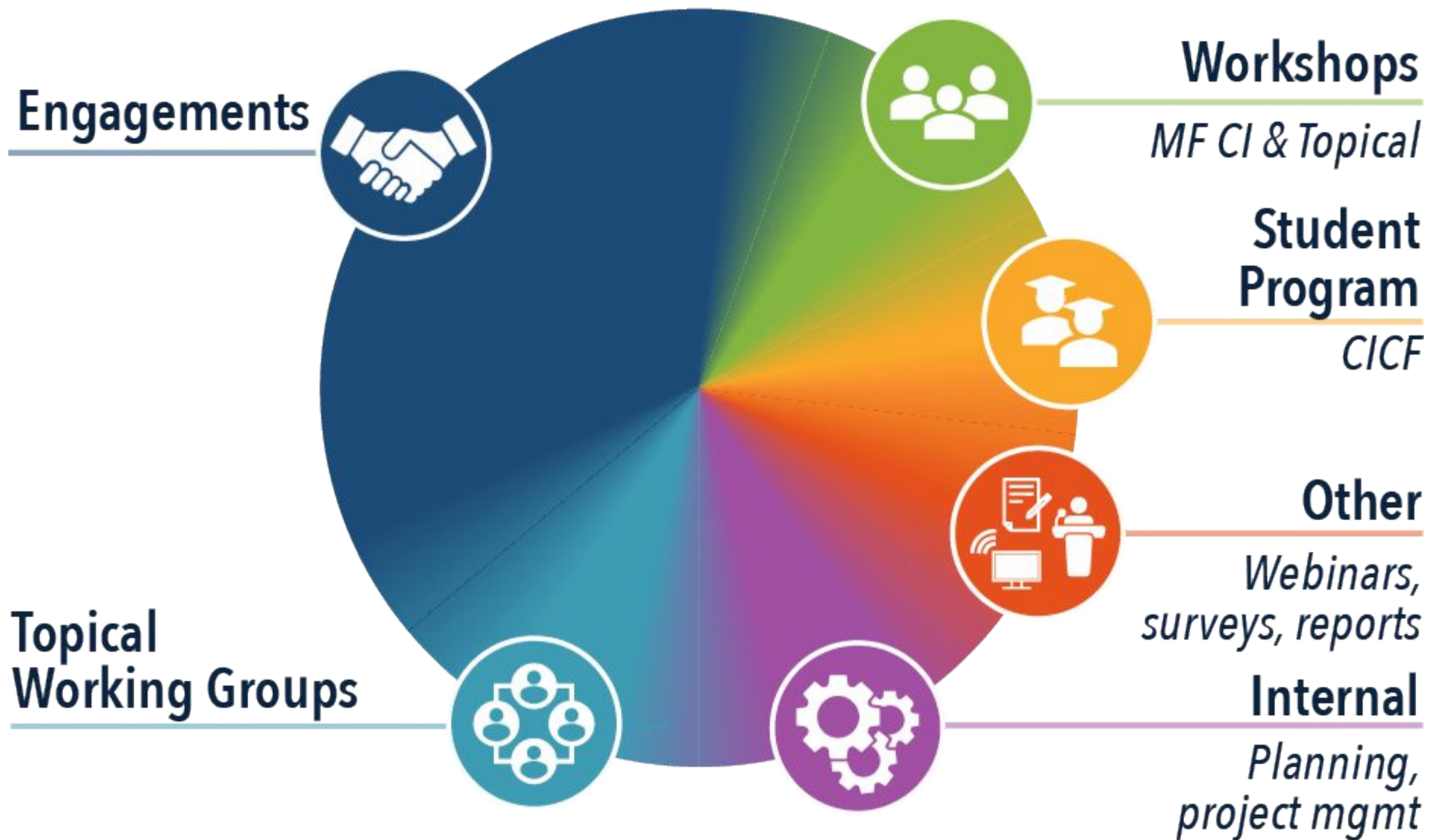


Mission



NSF CI Compass provides **expertise and active support to cyberinfrastructure practitioners at NSF Major Facilities in order to accelerate the data lifecycle** and ensure the integrity and effectiveness of the cyberinfrastructure upon which research and discovery depend.

CI Compass activities



CI Compass FAIR Topical WG

- Started in August 2022 to engage MFs to understand FAIR needs and share information



Data Lifecycle Questions/Problems

- How do facilities handle data from sensor to collection
 - metadata collected? capturing context
- Archiving and long-term management
 - Data Curation

Current Work

- Training
 - Data collection and deposit
- Community of Practice
 - MF Case Studies, presentations to TWG
- Facility Guidelines
 - Maturity Models!

- ***finite***: availability is expected to end on or around a given date (e.g., limited support for software versions not marked “long term stable”) or trigger event (e.g., single-use link).
- ***indefinite***: the provider has no particular commitment to the object.
- ***lifetime***: the object is expected to be available as long as the provider exists.
- ***subinfinite***: due to succession arrangements, the object is expected to be available beyond the provider organization’s lifetime.

**"Persistence Statements: Describing Digital Stickiness", J. Kunze, et al.
doi:10.5334/dsj-2017-039**

Identifiers

- Use persistent identifiers
 - ORCIDs
 - Things undoubtedly have an internal identifier, there is value in exposing that.
 - There is value even if identifier is not *resolvable*

Machine Interoperable Systems

No data is an island. Integration of data between systems is important.

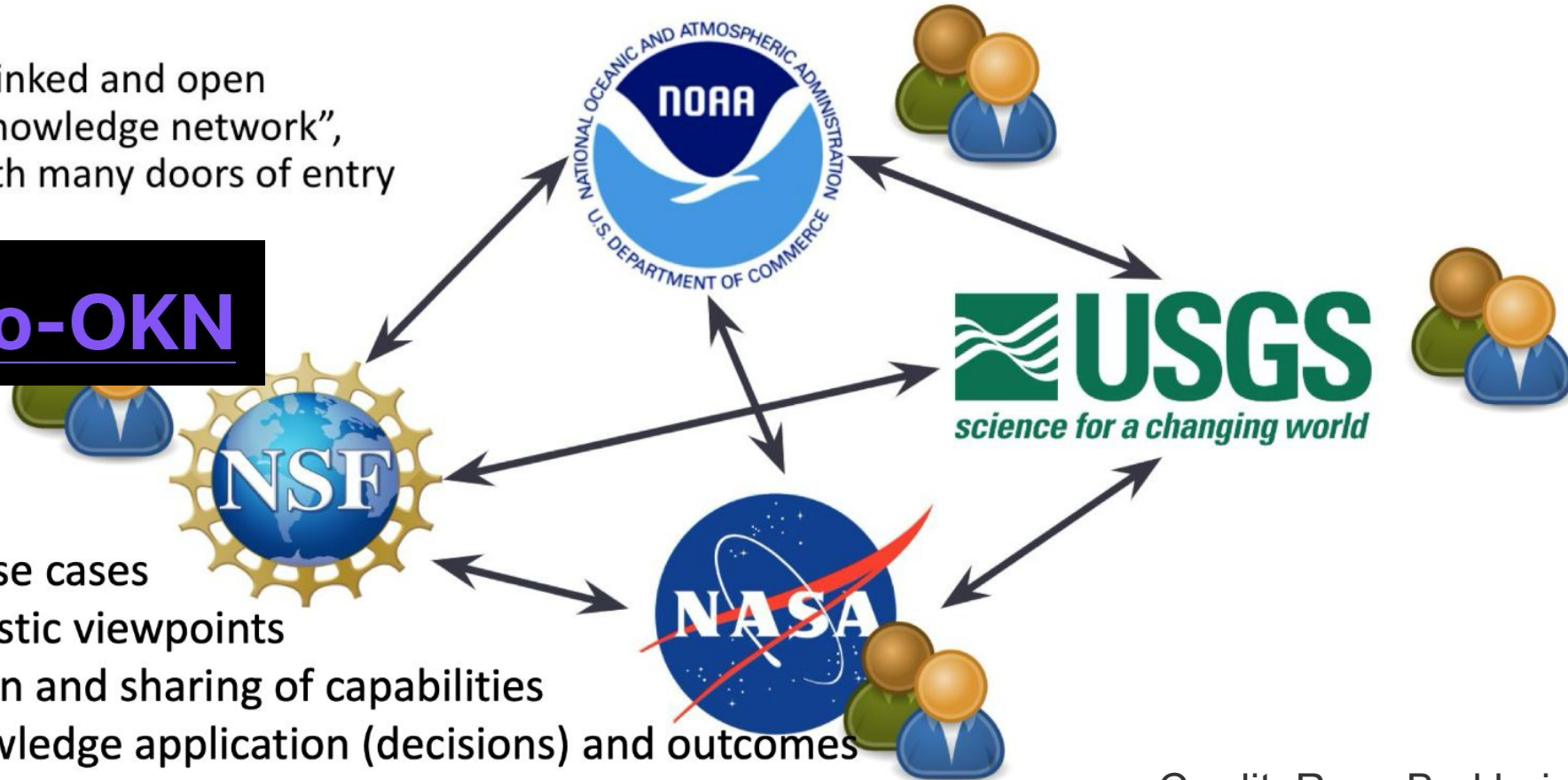
- Supports findability
- Cross-domain integration
- Provide Metadata in flexible systems.
- Knowledge Graphs
- Landing pages vs direct data access. [FAIR Signposting](#)



Toward an Open Ecosystem of Trusted Open Ecosystems

A linked and open "knowledge network", with many doors of entry

Proto-OKN



- more coverage in use cases
- providing more holistic viewpoints
- facility specialization and sharing of capabilities
- ability to track knowledge application (decisions) and outcomes

Credit: Ryan Berkheimer (NOAA/NCEI)

Levels of Identifiers

- Classes of things (RRID)
- Actual things (DOI)
- The minutiae (ARKs)
 - Provide a context and trail back to creator
 - Systems are already using a local identifier

AI Agent Ready Systems

How to build AI-ready data? Not only AI models, but AI Agents

- data is clean and well described
- data in a form that can be digestible to a LLM/RAG application
- use web standards such as Schema.org, JSON-LD, SPARQL

What are we trying to do???

advance science

steward previous and current work to future generations

discover how the world and universe works

improve our lives

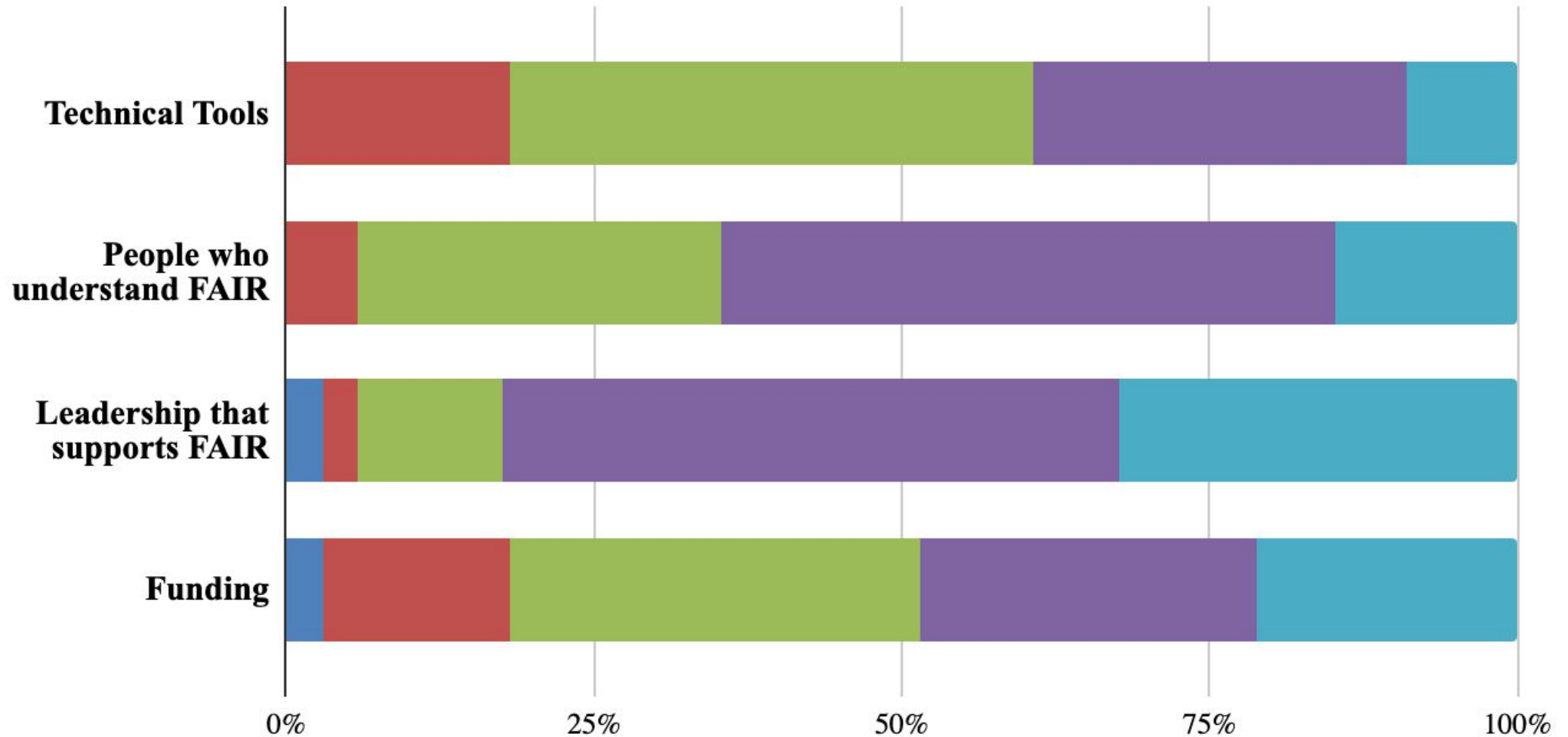
The wonder *and* joy of discovery

FAIR Survey

- Data collected March 2023
- 54 responses, with 9 follow-up interviews
- Goals
 - How do major facilities self-assess their FAIR practices?
 - Identify technologies being used for data management and FAIR

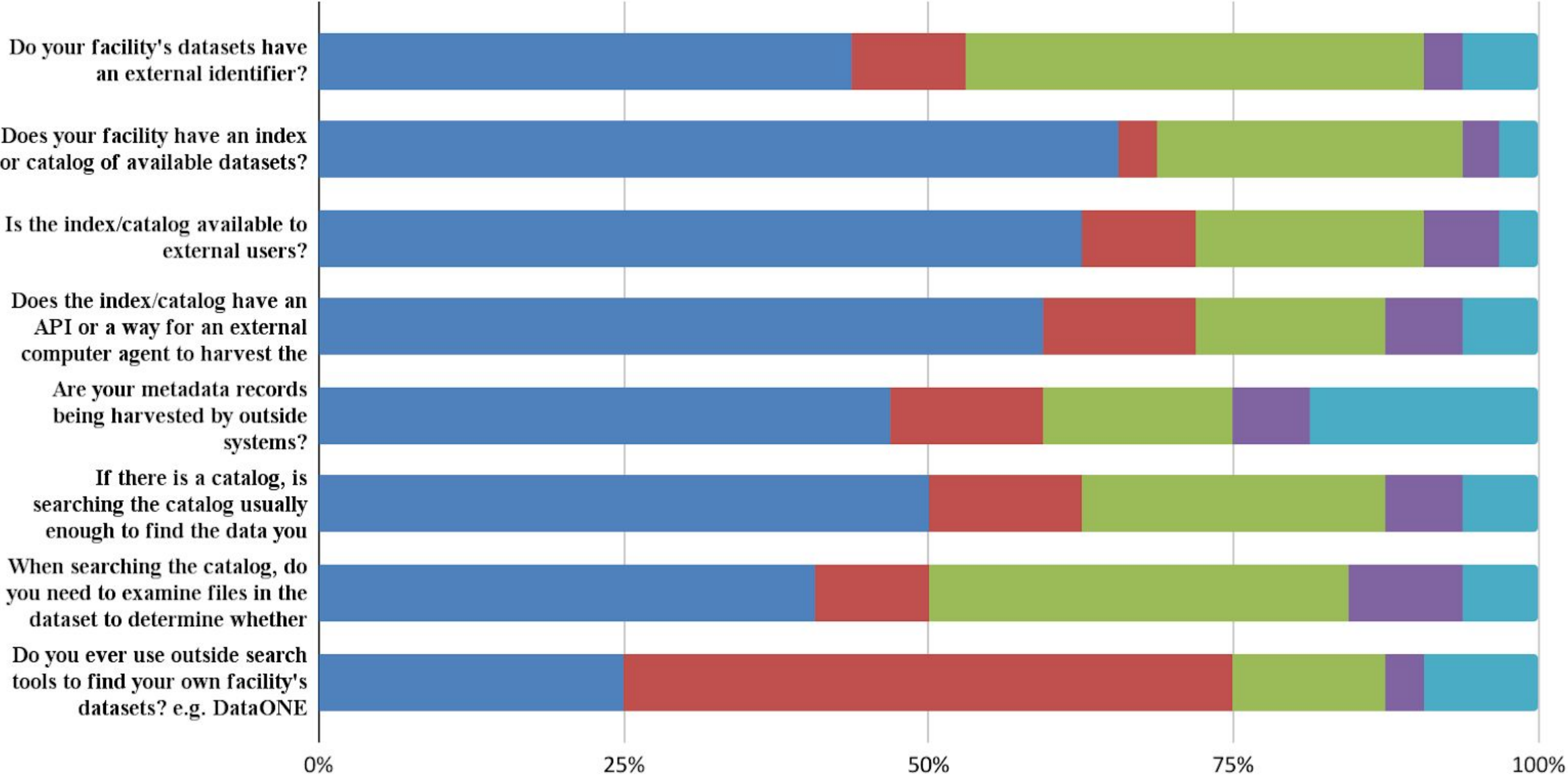
Which of the following have been helpful in regards to FAIR implementation at your organization?

■ Not at all helpful ■ Slightly helpful ■ Moderately helpful ■ Very helpful
■ Extremely helpful



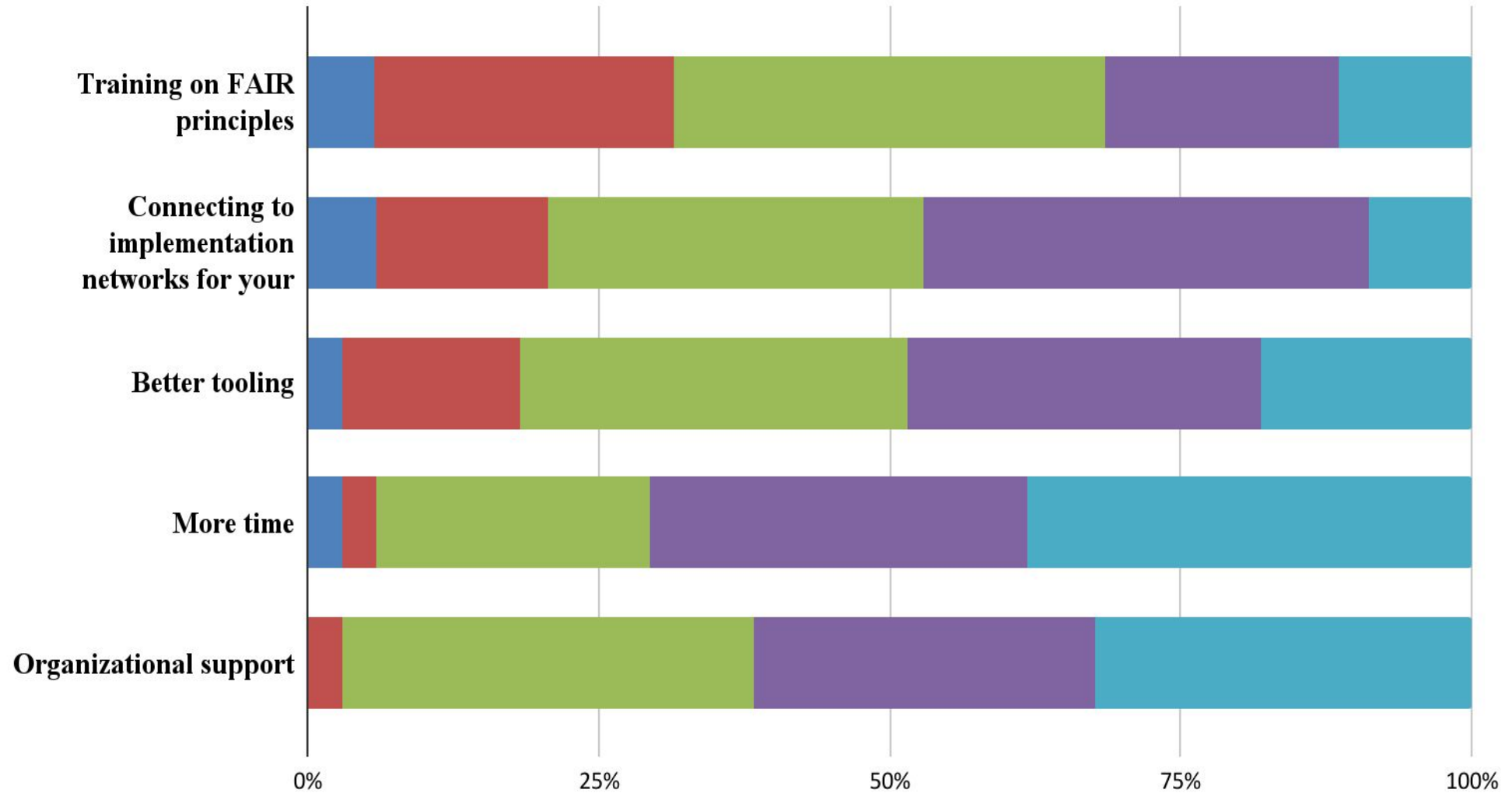
To the best of your knowledge does your facility do/have?

Yes No Some are Not Applicable I do not know



What would help you most in becoming more FAIR?

■ Not at all helpful ■ Slightly helpful ■ Moderately helpful ■ Very helpful ■ Extremely helpful



“Do you see AI/ML as a driver for FAIR data practices? If so, please describe.”

Yes	9	47%
No	4	21%
Unsure / Unrelated / Not Now	6	32%