MF Approach to Open Science: FAIR Data, Persistent Identifiers, and Engagement

David S. Butcher

2024 CI Compass CI4MF Workshop
Long Beach, CA

Headquartered at **Florida State University,** the MagLab also has branch campuses at the **University of Florida** and **Los Alamos National Laboratory.**

LANL
LOS ALAMOS, NM

FSU
TALLAHASSEE, FL

UF
GAINESVILLE, FL

MagLab at a Glance 2022 (https://nationalmaglab.org/about/print-media/at-a-glance)

MagLab Overview

NATIONAL MAGLAB

Headquartered at **Florida State University,** the MagLab also has branch campuses at the **University of Florida** and **Los Alamos National Laboratory.**

LANL
LOS ALAMOS, NM

FSU
TALLAHASSEE, FL

UF
GAINESVILLE, FL

- **Pulsed Field**
Short, ultra-powerful magnetic fields up to 100 T

- **High B/T**
Magnetic fields up to 15 T combined with ultra-cold temperatures of 0.4 mK

- **Advanced Magnetic Resonance Imaging & Spectroscopy (AMRIS)**
High-resolution solution and solid-state, NMR, animal imaging & human imaging

- **DC Field**
Steady, continuous magnetic fields up to 45 T

- **Electron Magnetic Resonance (EMR)**
Magnetic resonance techniques associated with the electron

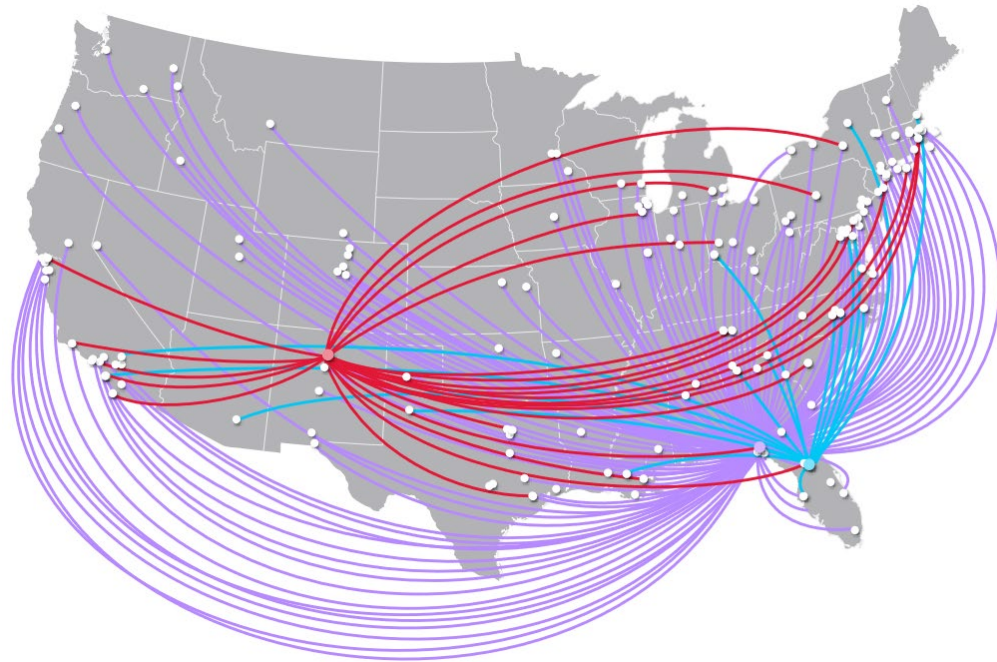- **Nuclear Magnetic Resonance (NMR)**
Solid & solution state NMR & animal imaging
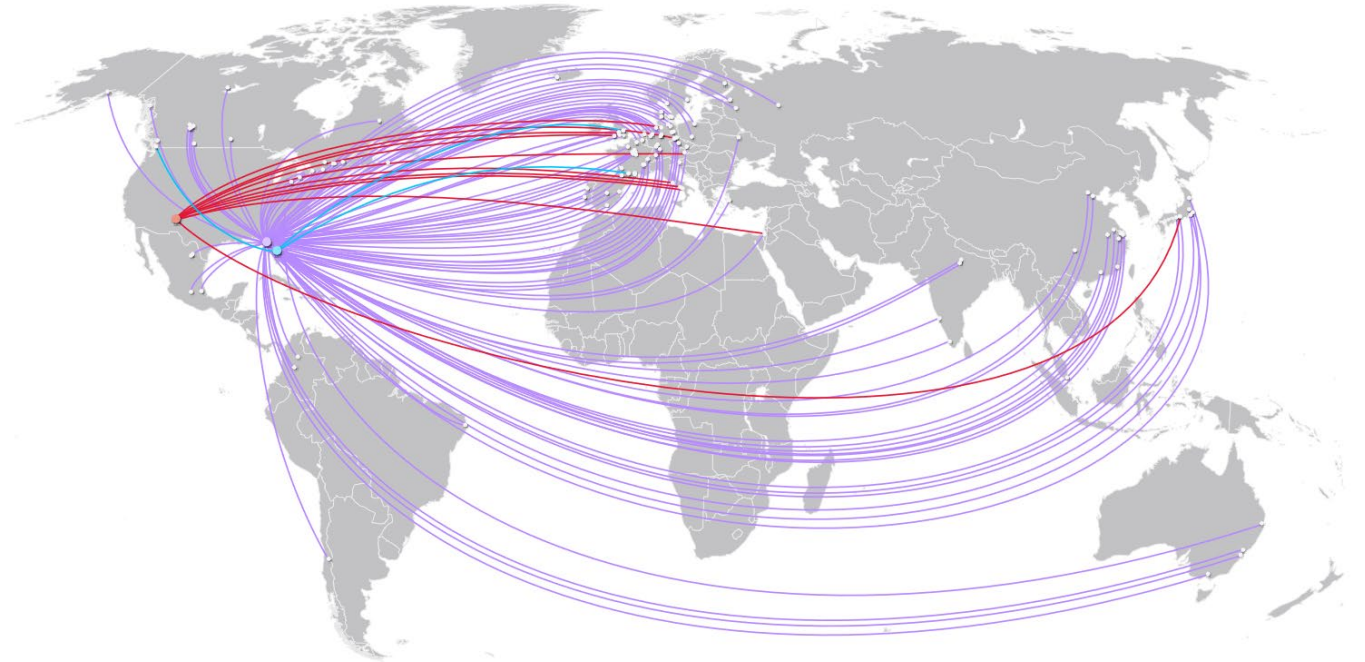
- **Ion Cyclotron Resonance (ICR)**
Ultra-high resolution and high mass accuracy Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry

In 2022, our **1958** users represented **327** universities, government labs and private companies worldwide.

In 2022, MagLab users published **352** articles in peer-reviewed journals.

MagLab at a Glance 2022 (https://nationalmaglab.org/about/print-media/at-a-glance)
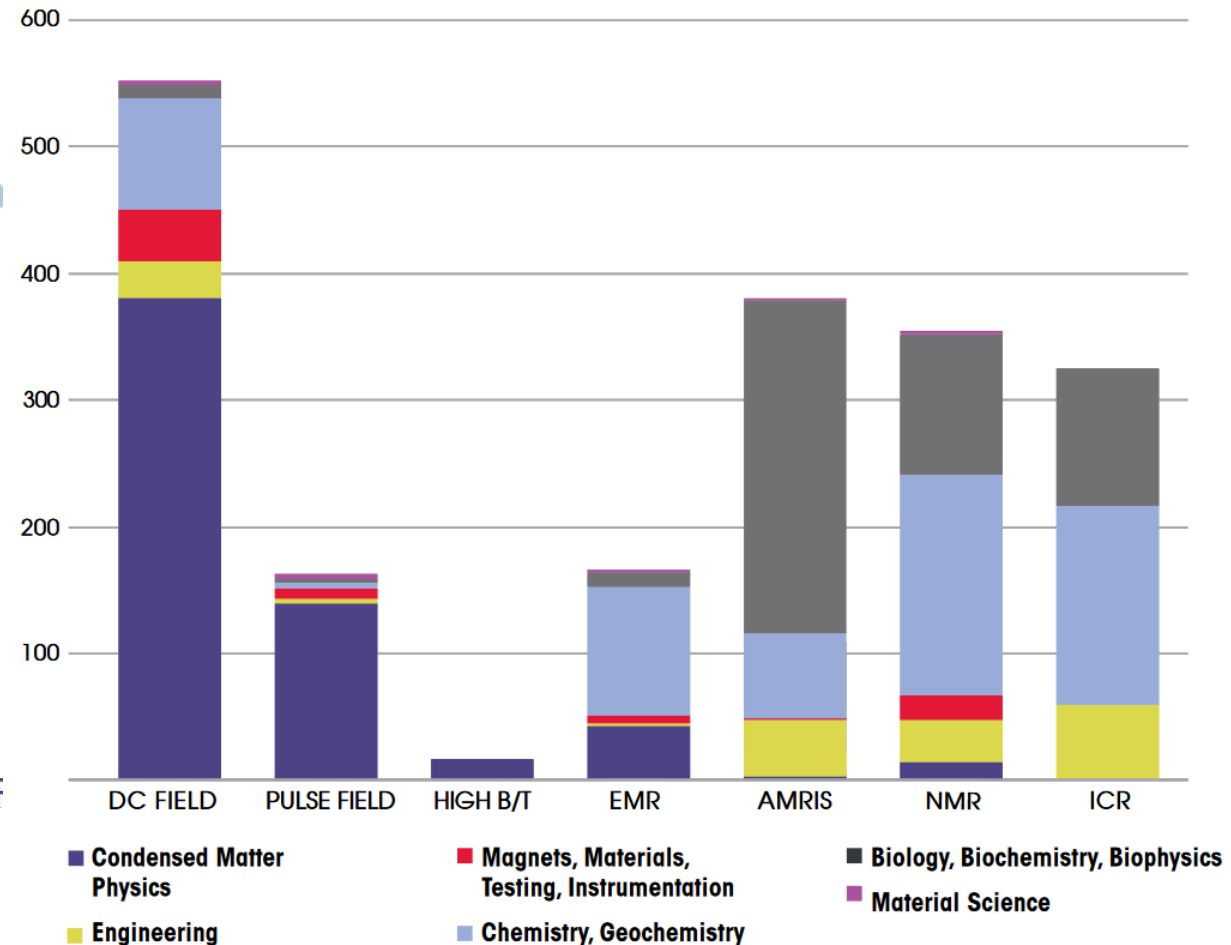
National MagLab

Headquartered at **Florida State University,** the MagLab also has branch campuses at the **University of Florida** and **Los Alamos National Laboratory.**

LANL
LOS ALAMOS, NM

FSU
TALLAHASSEE, FL

UF
GAINESVILLE, F

2022 USERS BY DISCIPLINE

Legend:
- Condensed Matter Physics
- Engineering
- Magnets, Materials, Testing, Instrumentation
- Chemistry, Geochemistry
- Biology, Biochemistry, Biophysics
- Material Science

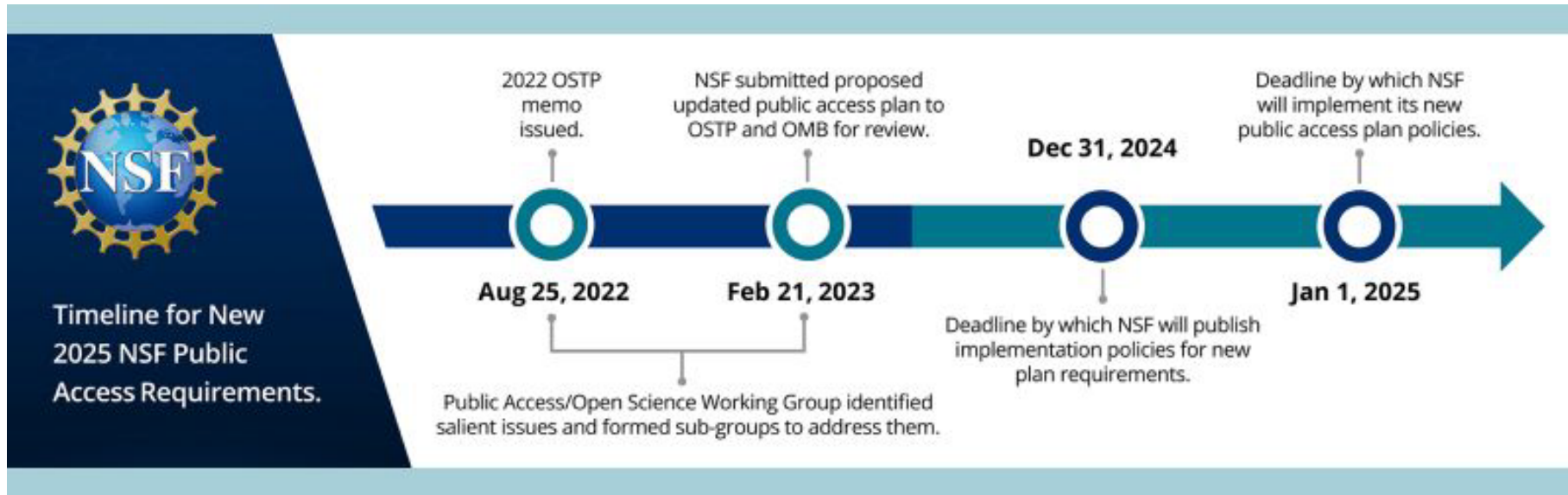MagLab at a Glance 2022 (https://nationalmaglab.org/about/print-media/at-a-glance)

The Meaning of FAIR & Open

The MagLab approach must consider a variety of factors…

- Facilitating and encouraging the desired behavior from diverse groups of users and analytical techniques

- Addressing the lack of a clear roadmap and established data sharing practices and appropriate CI in many disciplines

- Ensuring that funder requirements are met within upcoming deadlines

Timeline for New 2025 NSF Public Access Requirements.

2022 OSTP memo issued.

NSF submitted proposed updated public access plan to OSTP and OMB for review.

Deadline by which NSF will implement its new public access plan policies.

Dec 31, 2024

Aug 25, 2022

Feb 21, 2023

Jan 1, 2025

Deadline by which NSF will publish implementation policies for new plan requirements.

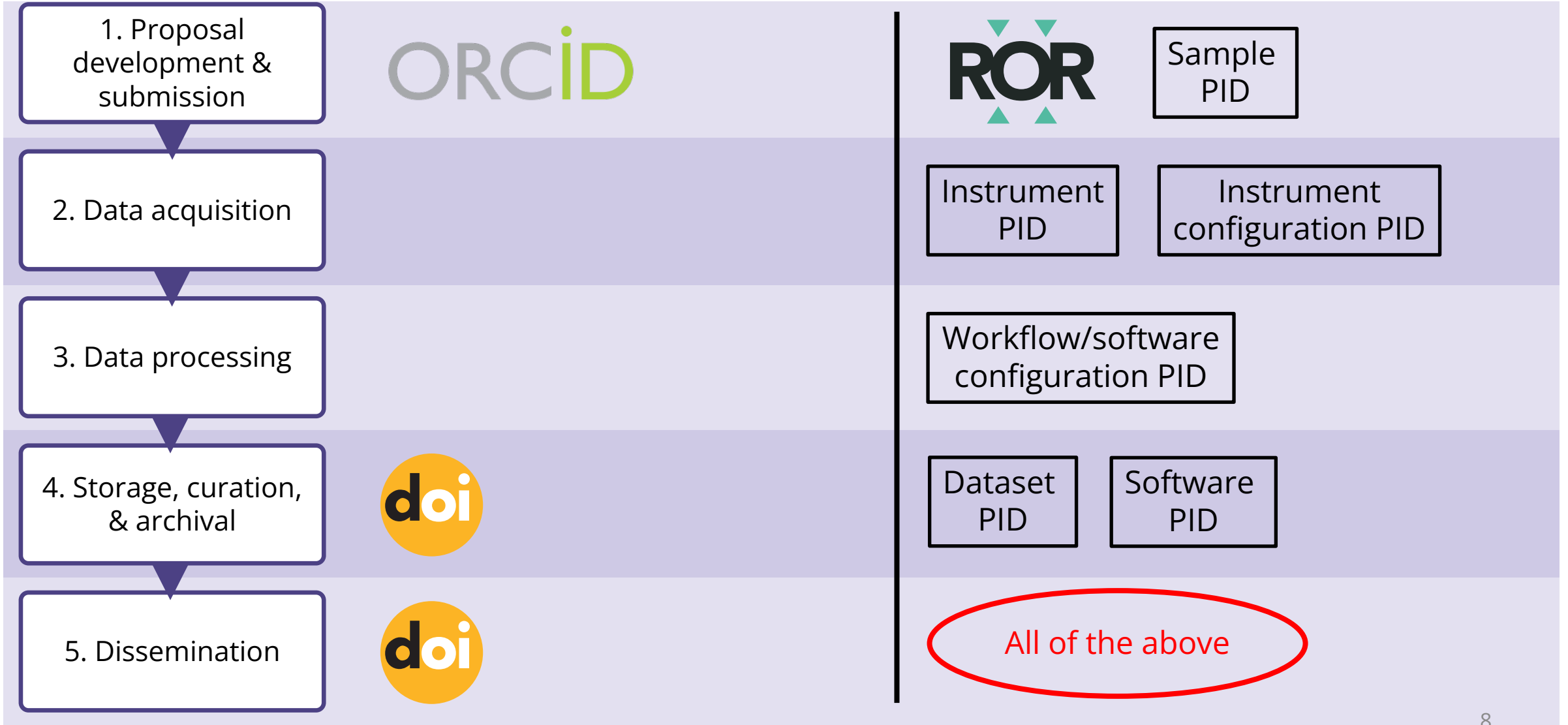Public Access/Open Science Working Group identified salient issues and formed sub-groups to address them.

"An update to the Public Access Plan focusing on researcher affiliations and funding and persistent digital identifiers for researchers, research products and awards will be published in 2026, for implementation by January 2027."[1]

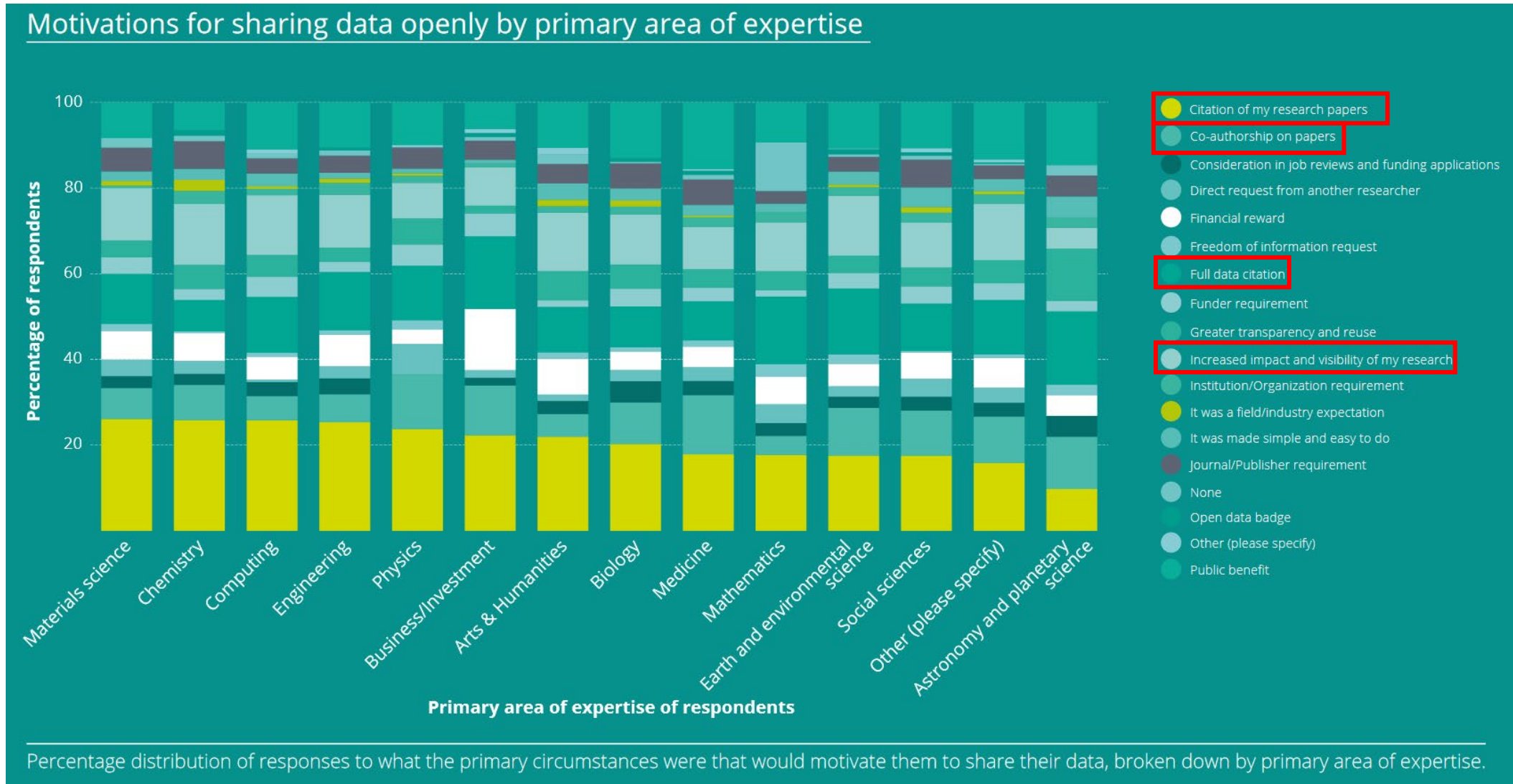1. NSF Public Access Initiative. *NSF - National Science Foundation* https://new.nsf.gov/public-access.

Implementation of PIDs

CURRENT

FUTURE

1. Proposal development & submission

ORCiD

ROR   Sample PID

2. Data acquisition

Instrument PID   Instrument configuration PID

3. Data processing

Workflow/software configuration PID

4. Storage, curation, & archival

doi

Dataset PID   Software PID

5. Dissemination

doi

All of the above

Driving Engagement

Motivations for sharing data openly by primary area of expertise

Percentage distribution of responses to what the primary circumstances were that would motivate them to share their data, broken down by primary area of expertise.

Driving Engagement

Driving Engagement

2022 Internal MagLab FAIROS Survey

Do you feel that the application of FAIROS principles (*e.g.* making data available for reuse) is important for products of research at your facility? (119 responses)
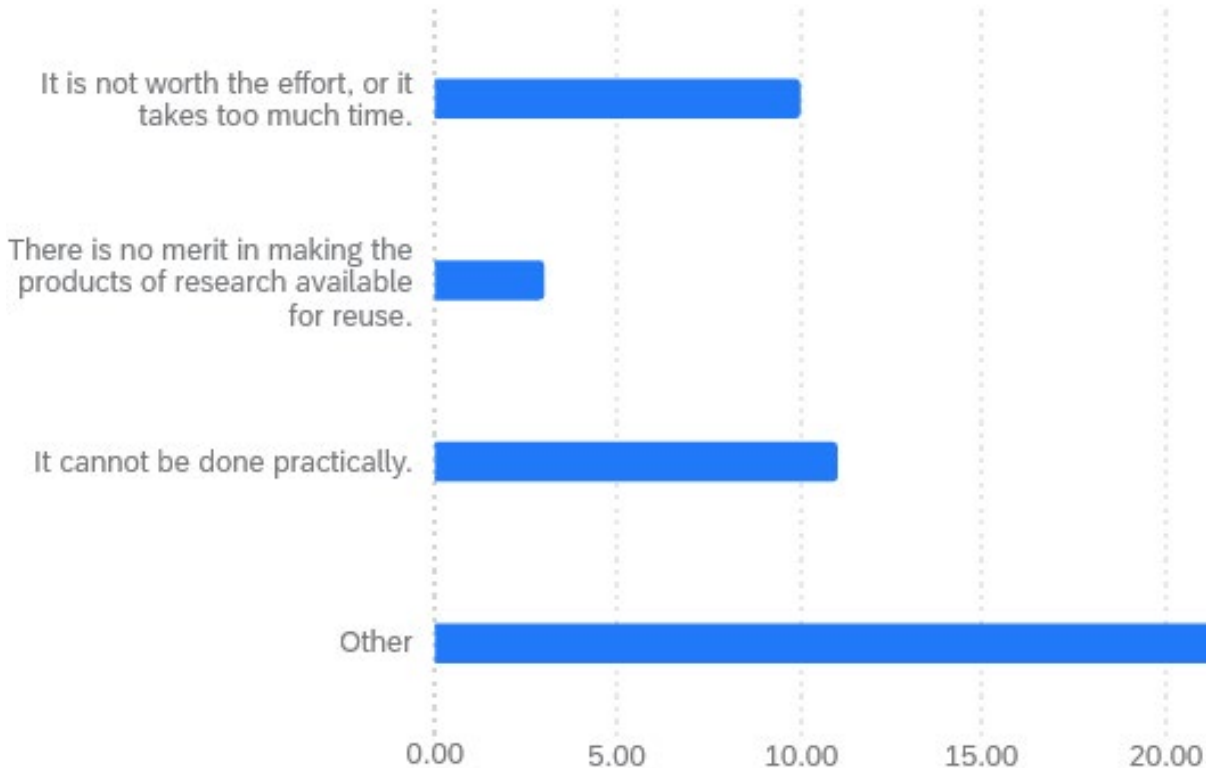

Do you feel that the application of FAIROS principles (e.g. making data available for reuse) is important ... ⓘ
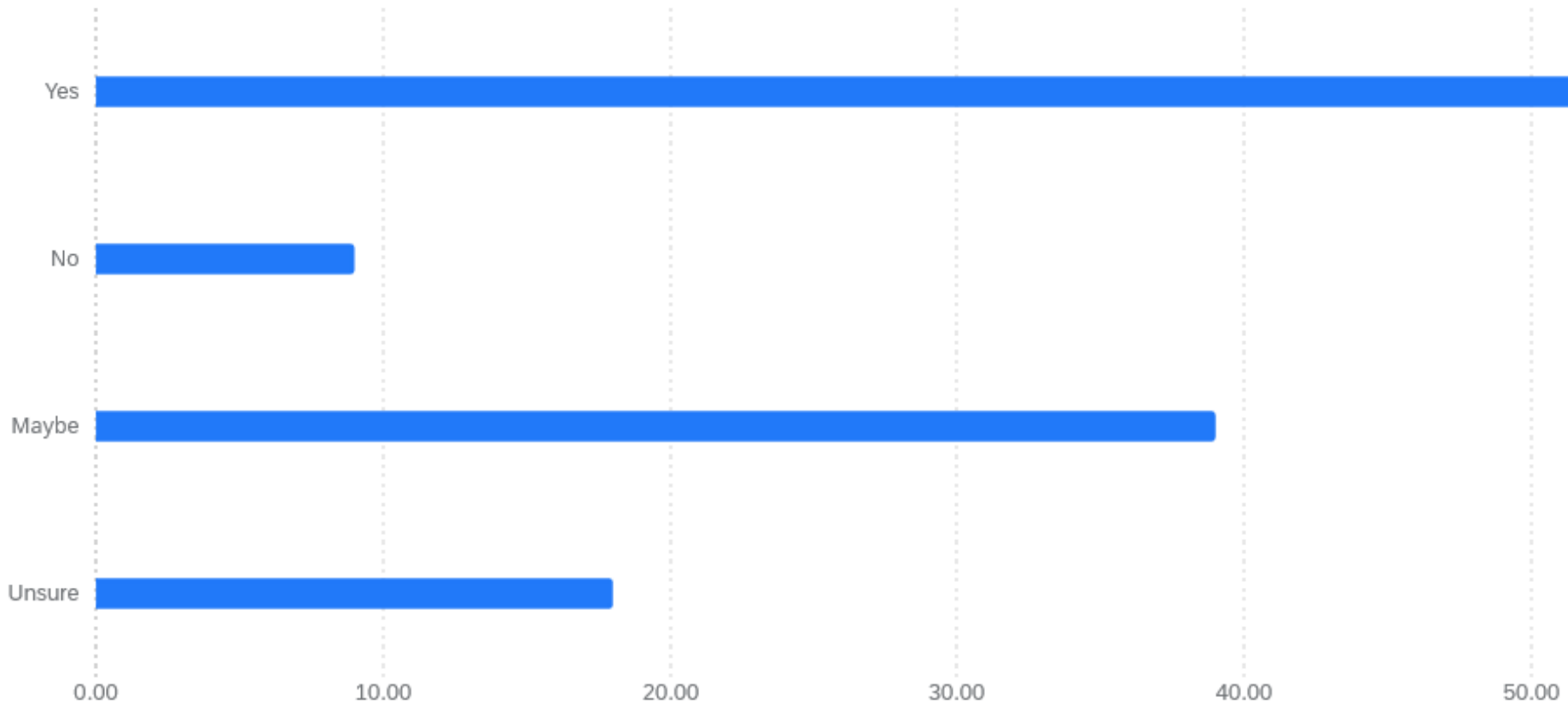
Driving Engagement

2022 Internal MagLab FAIROS Survey

Why not? ⓘ

Selected "Other" Responses

"I am concerned about scientists' ability to perform experiments without immediate plans to publish the data."

"I haven't seen a good example of it being used in my field of research."
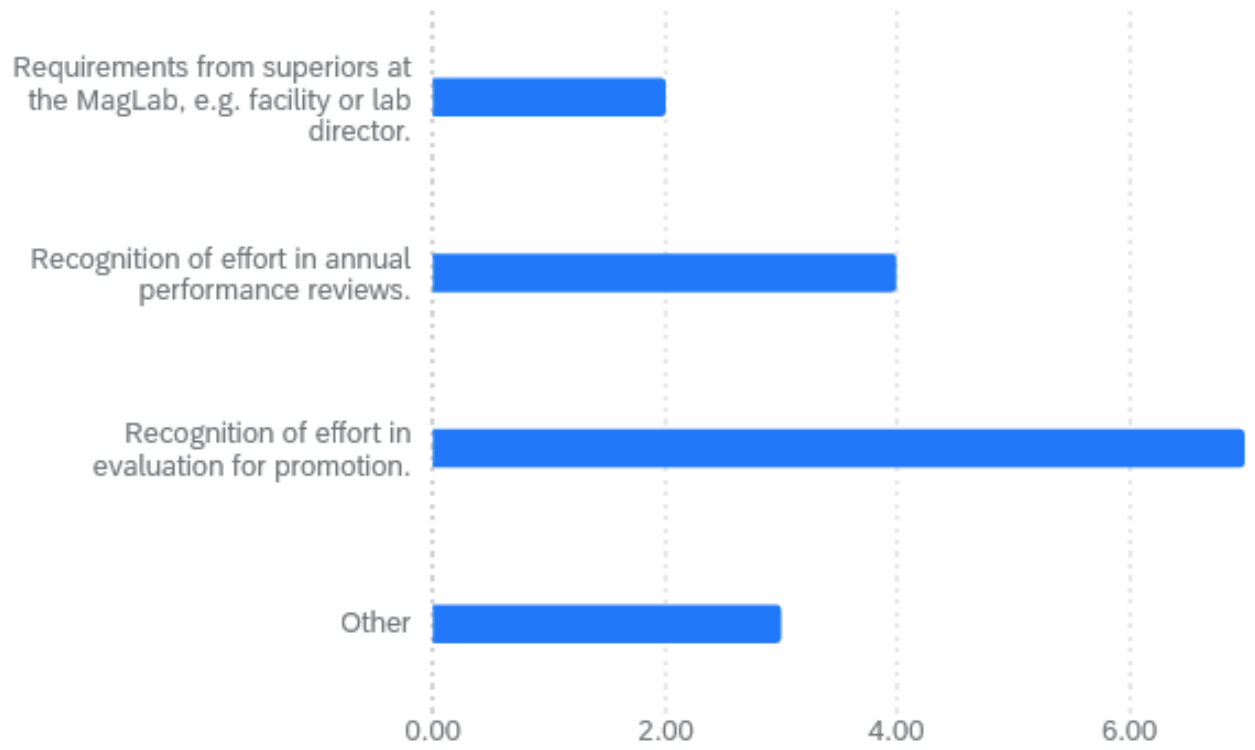
"I am worried that my data will be misused by a competitor."

"There is no interest/demand for it in my field of research."

Driving Engagement

2022 Internal MagLab FAIROS Survey

What incentives would make it worth the time/effort? (10 responses)

Selected Responses

"An annual or quarterly award."

"None of these are of interest to top rank faculty."

"None of the above."

MagLab staff are often not aware of appropriate incentives for making data FAIR and open.

Survey results suggest (unsurprisingly) that we can drive engagement by <u>appealing to researchers' self interest.</u>

Driving appropriate use of CI

Future Concerns

- Will NSF's funding strategy or roadmaps for researchers change?
  - Funding for cyberinfrastructure, RCNs, etc.
  - Funding for user facilities is provided through a variety of different mechanisms – how will data sharing needs be incorporated?
- Will publishing models change in the future? How, and how quickly? Will those changes help to accelerate an adoption of FAIR and open science?
  - Plan S for pubs focuses on being community-driven – will community want to drive in the "right" direction?
- AI built on FAIR and open data will become a part of the scientific process. How can we ensure it is being used responsibly?
  - Already evidence of use of AI LLMs for papers – evidence is purely anecdotal
  - Do MFs have any role in regulating or facilitating the use of AI?

Persistent Identifiers

- PIDs can be a driver of change
  - Good value proposition – its useful that every gene in human genome has a unique identifiers
 - Can't be

Facilitating Reuse

Small Steps

- The engagement gap – how do we get researchers/data creators to actually do something if there is no clear roadmap?
- Build specific activities that facilitate FAIR into existing processes (e.g. publication). For new processes (Plan S), make sure they are built in from the beginning
- Evolutionary and/or revolutionary change
- Education of students

Enabling users to meet data sharing requirements
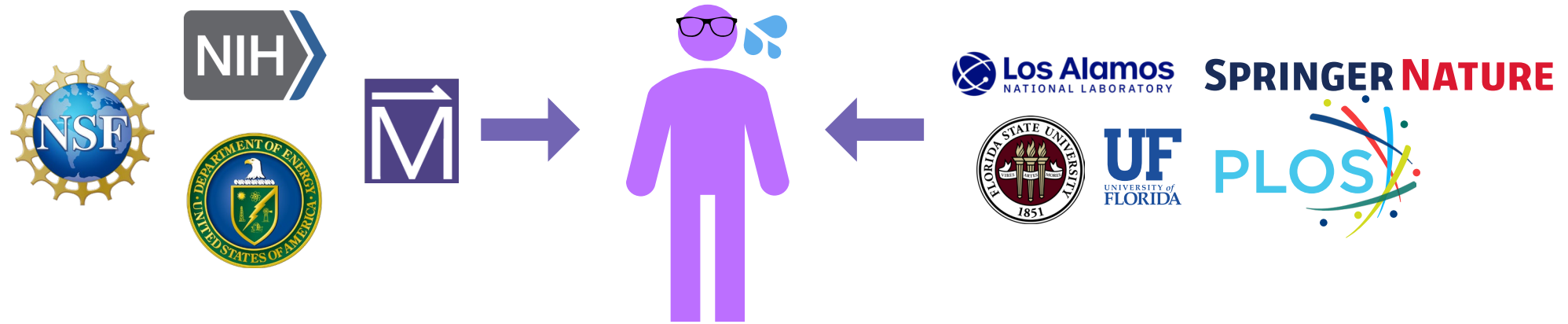
Ideal

- Automated capture and packaging of data and metadata with PIDs for vocabulary and context
- Automated upload to an appropriate repository which assigns PID and facilitates reporting to publishers and funders

Challenges

- Current procedures require extensive manual input for a well-annotated product
- Different funding agencies/international users not part of the system
- Varying policies between funders/publishers

PID Use Case 2

Facilitating reuse by data users
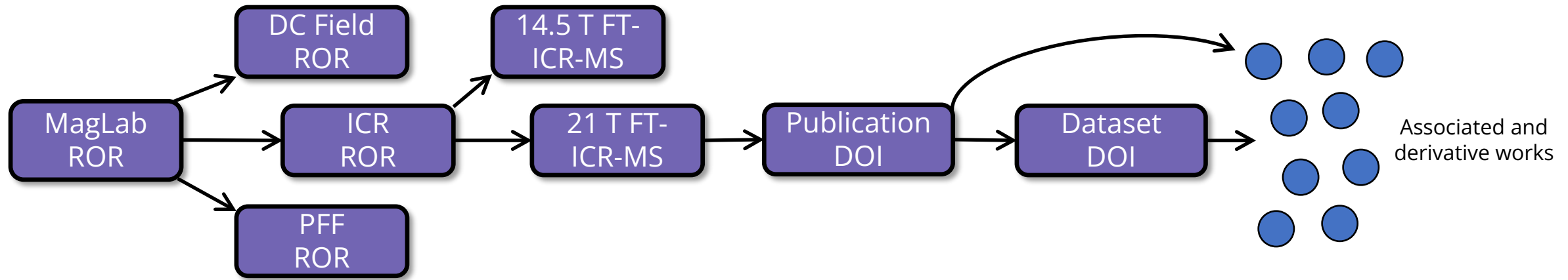

Ideal	Challenges

Ideal

- Complete sample metadata
- Detailed and accessible instrumental metadata, including both stable and changing components
- Replicable analysis workflow
- Complete and accessible datasets and publications

Challenges

- Radically different sample types
- Instrumental setups difficult to capture in detail
- Administrative overhead from PID assignments and maintenance
- Gaps in user knowledge/unavailable information

Tracking the creation and propagation of products of MagLab research

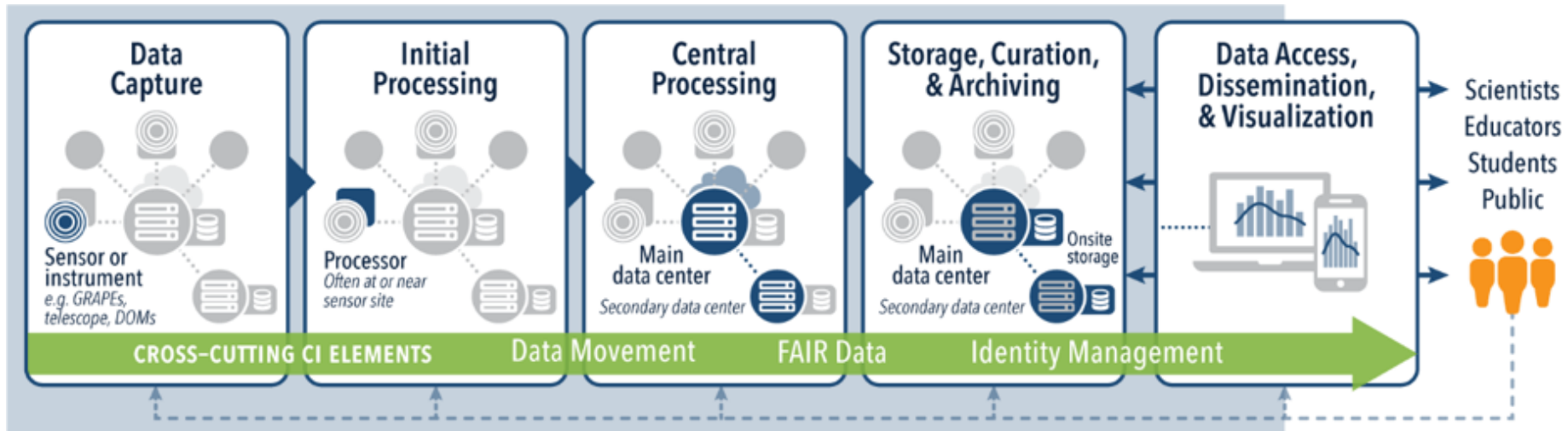| MagLab ROR → DC Field ROR |
| MagLab ROR → ICR ROR → 14.5 T FT-ICR-MS |
| ICR ROR → 21 T FT-ICR-MS → Publication DOI → Dataset DOI → Associated and derivative works |
| MagLab ROR → PFF ROR |

Ideal

- Assessment of the MagLab's impact on the body of published literature – both direct and indirect.
- Comparative analyses with other facilities
- Tracking connections between users and colleagues and recruiting new users
- Tracking global reach

Challenges

- Missing extremely relevant PIDs
- Missing a way to assign them and have them associated with the data product
- Existing/in-development standards may not be suitable for all disciplines
- User reporting is extremely inconsistent/incomplete

CI Compass Data Lifecycle Model[1]

MagLab Data Lifecycle Model

1. The Major Facilities Data Lifecycle in a Nutshell. https://doi.org/10.5281/zenodo.5550224