FAIR Digital Objects in Distributed Research Environments

Throughput Computing July 11 2023

Natalie Meyers
nmeyers@sdsc.edu

This work funded by the National Science Foundation, award 2138811

NSDF National Science Data Fabric

SDSC SAN DIEGO SUPERCOMPUTER CENTER

Advancing FAIR in the US

The GO FAIR US Support and Coordination Office is hosted by the San Diego Supercomputer Center (SDSC), University of California San Diego.

www.gofair.us

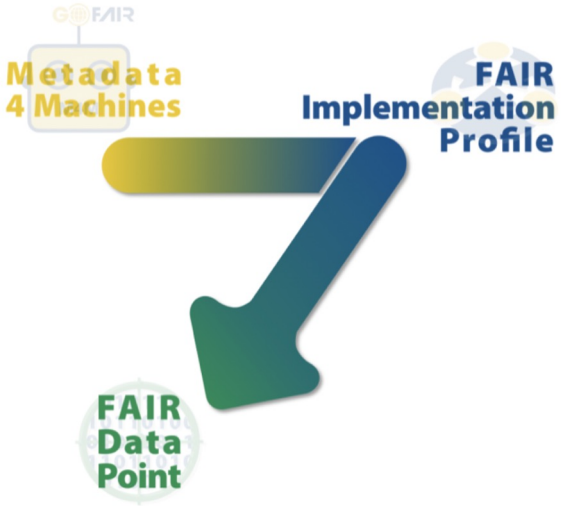
Portions of this GO FAIR US work supported through the NSF awards #1928208 and 2226453, and SDSC seed funding.

National Science Data Fabric

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

Unique Aspects of GO FAIR US

- Ambassadors Pillar - headed by Nancy Hoebelheinrich
  - Responsive to gap of estimated 100s of 1,000s of data stewards needed
  - Scaling up consultants and capacity affiliated but outside of GO FAIR
  - Leads for domains (Earth Sciences, Materials Science)
- Focus on FAIR implementation
  - Practical solutions in US funding environment
  - Scanning developments outside the US and translating back
  - Resources and expertise for piloting approaches
- Sector inclusive - GO FAIR US leadership includes
  - Industry, small business
  - Researchers funded by NSF, DOE, NIH
  - Funder

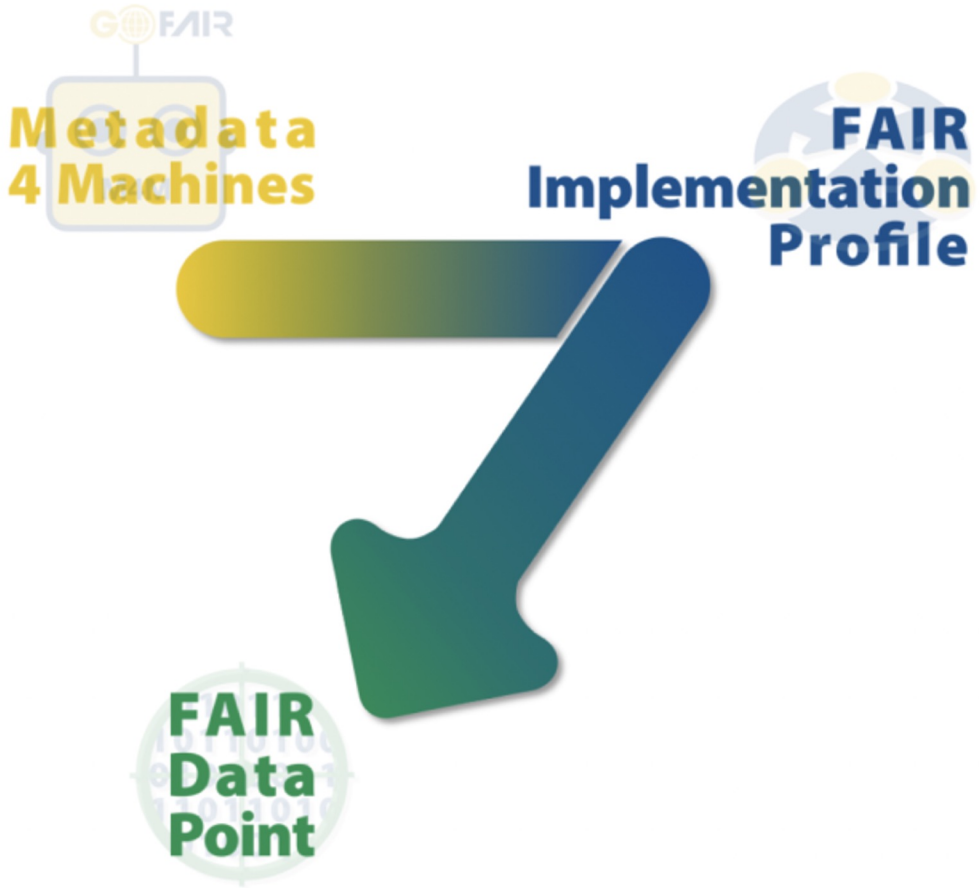


GO FAIR US 2023

- Outreach, consultations, network building, webinars, blog posts
  - Focus on connecting with early career via FAIRPoints
- Effort following funded activities
  - FARR - FAIR in ML, AI Readiness, & Reproducibility
    - FAIR OS RCN, Farr-rcn.org
    - WDS connection serving & collaborating with Geoscience repos
    - Working through RDA IG and ESIP clusters
  - SeekCommons.org - Socio-environmental Knowledge Commons Project
  - National Data Science Fabric
    - Lightweight cataloguing & metadata extraction
    - Piloting FDO specification
- Life sciences focus
  - Adapting and scaling M4M workshops from lessons learned in M4M for Geosciences
- Connecting M4M and FIP work to CODATA/RDA's WorldFAIR and WorldFAIR+, seeking more US petals
- FAIRIST

FARR
FAIR, AI Readiness & Reproducibility

Three-Point FAIRification



A systematic process for FAIRification that can empower convergence

GO FAIR US

FAIR Implementation Goals for NSDF

Vision: NSDF as a FAIR-enabling Platform

- NSDF allows for easy adoption of FAIR community practices
- FAIR components part of the NSDF value proposition
- First large-scale infrastructure in the US to use FAIR Digital Objects and other novel approaches
- OSG as a FAIR enabling Partner

FAIR Principles

Findable

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. metadata specify the data identifier.
F4. (meta)data are registered or indexed in a searchable resource.

Interoperable

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

Accessible

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

Reusable

R1. meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards.

Wilkinson, 2016. https://doi.org/10.1038/sdata.2016.18

Findable Goals

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. metadata specify the data identifier.
F4. (meta)data are registered or indexed in a searchable resource.

- Ensuring objects have associated PIDs
- Applications and processes include capability for adding metadata & provenance
- Data is registered/deposited in a repository
- NSDF UIs include schema.org tags
- NSDF catalogs exist, provide useful searches for use cases
- NSDF provides on ramps to taking first or next steps with improved metadata and related standards

Accessible Goals

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

- Query metadata via an API
- Access is open and free

Interoperable Goals

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

- Metadata is machine readable (JSON LD, RDF)
- Partners increasingly use standard vocabularies (that embody the FAIR Principles)
- Metadata includes reference to standard and facilitates cross walks

Reusable Goals

R1. meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards.

- NSDF portals provide easy way to include license (CC-BY, etc.)
- Provenance capability
- Metadata uses standard listed on fairsharing.org

NSDF Activities Contributing to FAIR Goals

Current Work

Further catalog work

- Enables findability
- Stimulates work on metadata
- Catalogue OSG, OSN
- Provide Dataverse or Clowder catalog alongside NSDF catalogue for comparison
- Advertise data that can be visited (by machines)

Future Work

- Integrating Cross Domain Interoperability Framework (CDIF), schema.org & extensions, FDO specifications, and DeCODER
- Working with use cases to implement ORCIDs, ORCID API for login

NSDF Activities Contributing to FAIR Goals

Current Work

Further catalog work

- Enables findability
- Stimulates work on metadata
- Catalogue OSG, OSN
- Provide Dataverse or Clowder catalog alongside NSDF catalogue for comparison
- Advertise data that can be visited (by machines)

Future Work

- Integrating Cross Domain Interoperability Framework (CDIF), schema.org & extensions, FDO specifications, and DeCODER
- Working with use cases to implement ORCIDs, ORCID API for login

Thank You

Contact:
Natalie Meyers
nmeyers@sdsc.edu

www.gofair.us