



FAIR Basics

Juliane Schneider
Senior Bioinformatics Analyst/Data Liaison
Sage Bionetworks

“The FAIR Data Principles are oblique and faintly annoying”

-- Ada Lovelace



FAIR is...

A set of principles that describe the attributes data need to have to enable and enhance reuse, by humans and machines.

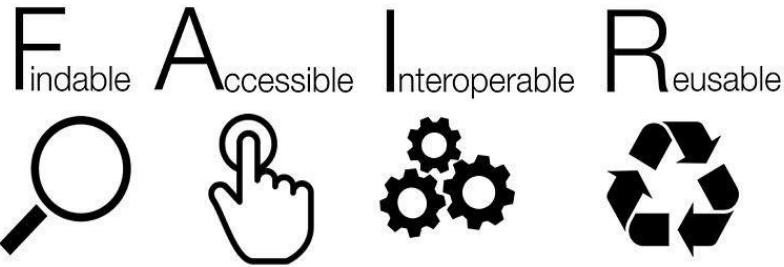


Image CC-BY-SA by [Sangya Pundir](#)

Where to find them and examples of use

The FAIR Data Principles

(<https://www.force11.org/group/fairgroup/fairprinciples>)

Seminal paper on them (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC479217/>)

A step by step how-to on using FAIR, in this case how PHI-base used the principles to publish FAIR data:

(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4922217/>)

Things FAIR cares about

Metadata

Identifiers

Machine-readability

Indexed information

Communication standards

Clear, stated usage rules (aka, a license)

Things FAIR does not care about

Disciplines/domains

Point-for-point application of the principles

Meeting every single point, relevant or not

Using a specific tool or method

Human beings

Findable and Accessible do
NOT mean open access

Not all data is meant to be public

FERPA (students)

HIPAA (medical)

FAIR PRINCIPLES

Findable:

- F1 (meta)data are assigned a globally unique and persistent identifier;
- F2 data are described with rich metadata;
- F3. metadata clearly and explicitly include the identifier of the data it describes;
- 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:

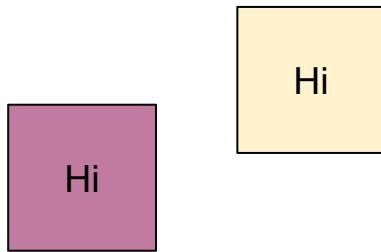
- R1. (meta)data are richly described with a plurality of accurate and relevant attributes;
 - R1.1 (meta)data are released with a clear and accessible data usage license;
 - R1.2 (meta)data are associated with detailed provenance;
 - R1.3 (meta)data meet domain-relevant community standards;

DOI 10.1038/sdata.2016.18

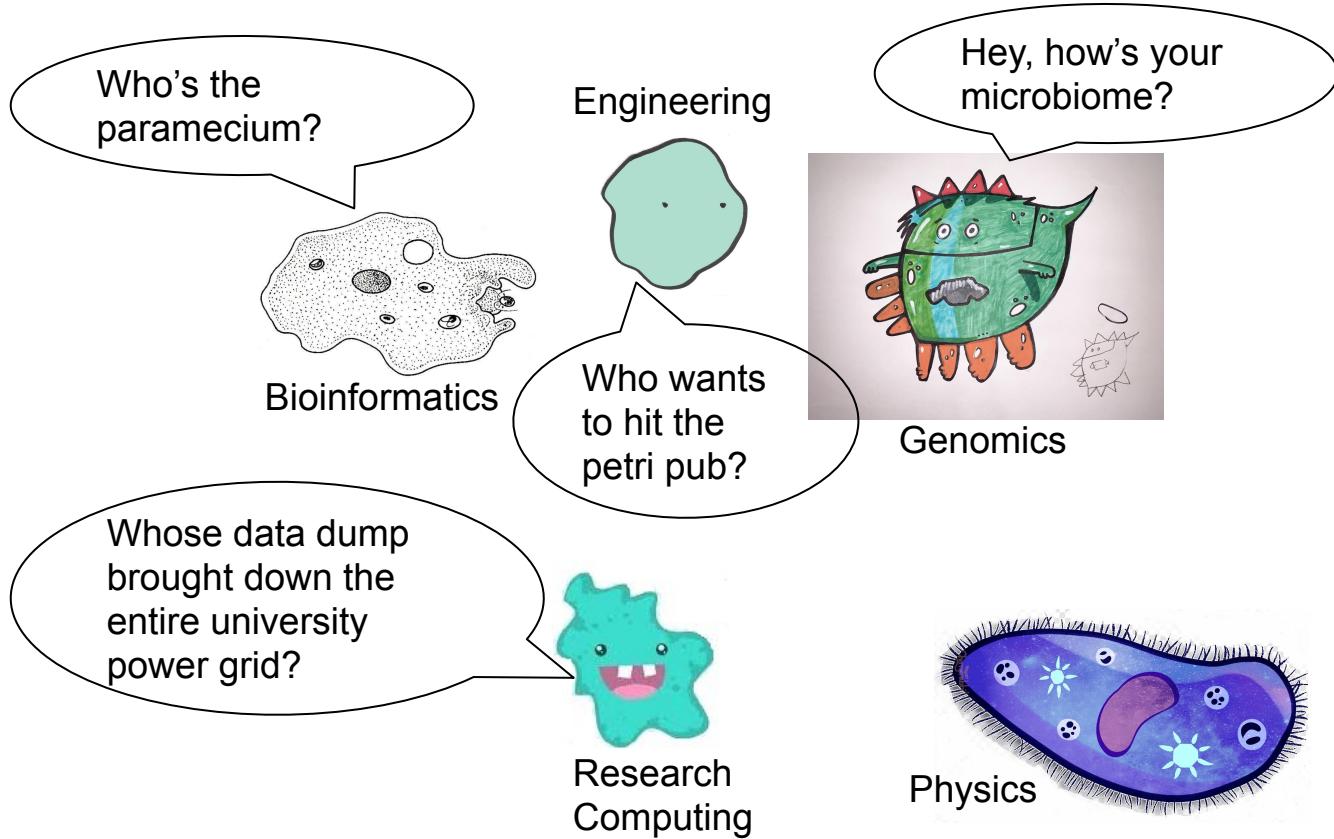
Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016).

FAIR Principles references “community standards”

Community is not a box



More like a super social amoeba



For Example, Bioinformatics work with:

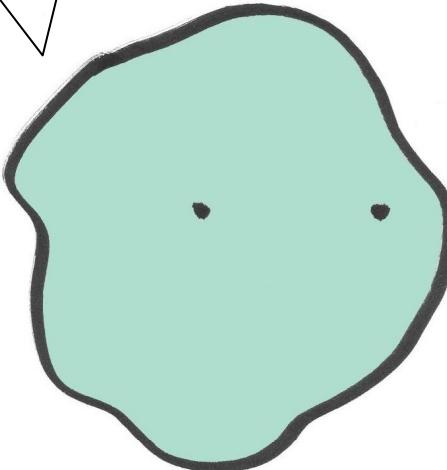
- Clinicians - EHR data
- Linguists - ontological data/algorithmic data
- Biologists - Ontology for Biomedical Investigation
- Geneticists - Gene Ontology
- Computer Scientists - software and algorithms
- Data Curators - everything
- Grant Administrators - data use agreements, funder requirements

Which do you choose?

CHOOSE ME
CHOOSE ME
CHOOSE ME



whatever



It depends

*FAIR is a tool to use to find
solutions for sharing your
data in the way you need it
to be shared*



Convergence

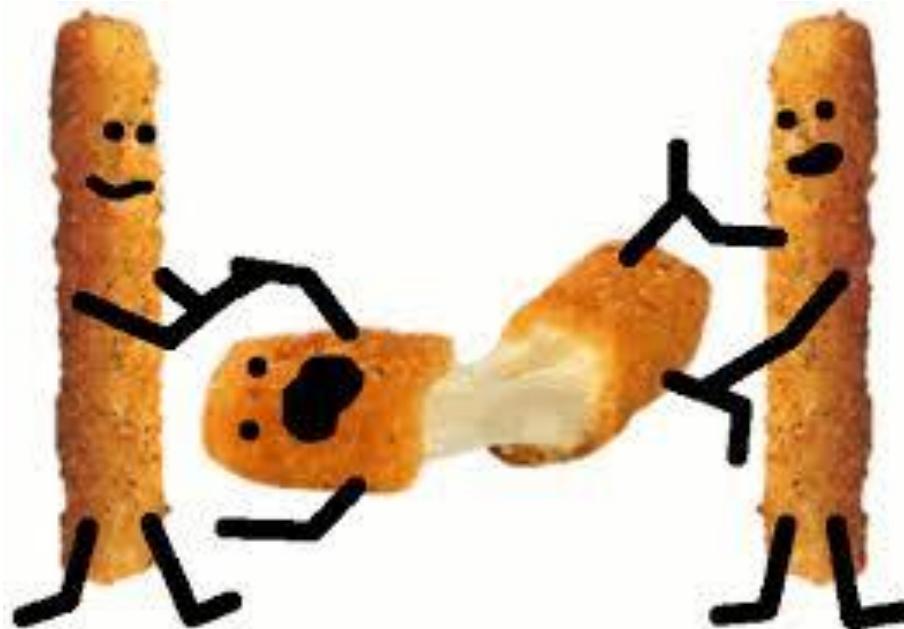
Symposium Nov. 27-Dec. 4

<https://www.go-fair.org/events/international-fair-convergence-symposium/>

The International FAIR Convergence Symposium will provide a **forum for advancing international and cross-domain convergence around FAIR**. The event will bring together a global data community with an interest in combining data across domains for a host of research issues

Meanwhile....

FAIR is a team sport



Sources of Truth - Findability

- Is the metadata indexed anywhere, does it have a persistent identifier (PID) and does it reference the PID of the data it describes?
 - Librarians, the repository team, or the IT group
- Is the data described with rich metadata?
 - Creator of the data (lab, person, PI) and/or the librarian who works on describing it

Sources of Truth - Accessibility

- (meta)data are retrievable by their identifier using a standardized communications protocol that is open, free, and universally implementable.
 - Repository group, IT department
- the protocol allows for an authentication and authorization procedure, where necessary.
 - Repository group, librarians, data curators
- metadata are accessible, even when the data are no longer available.
 - Librarians, data curators

Sources of Truth - Interoperability

- (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
 - Data curators, librarians, repository group
- (meta)data use vocabularies that follow FAIR principles.
 - Librarians, data curators
- (meta)data include qualified references to other (meta)data
 - Librarians, data curators

Sources of Truth - Reusability

- (meta)data are released with a clear and accessible data usage license.
 - Repository group, librarians, data curators
- (meta)data are associated with their provenance.
 - Repository group, librarians, data curators
- (meta)data meet domain-relevant community standards.
 - Data creator, librarians, data curators

But can you actually use it for anything?

WH



What do you use it for?

- To meet funder requirements for data
- To make your data discoverable to your colleagues, potential collaborators
- To build tools and technologies that support FAIR Data Principles
- To build repositories that allow for data that are findable, interoperable, accessible, reusable
- To evaluate a current repository for those same things
- To stymie administrators and/or get invited to present on FAIR

Funder requirements you may see:

- Your data needs to:
 - Be uploaded to an open access repository
 - Open access does NOT mean ability to see the data, but to find it
 - Data should conform to standards
 - Dates (ISO 8601)
 - Geospatial metadata (ISO 19115)
 - Data should be described by metadata, preferably using some kind of ontology or standard
 - Have PIDs

GO-FAIR Fair Funder Cycle

<https://www.go-fair.org/today/FAIR-funder/>

“international stakeholders (including two European research funding agencies) conceived the FAIR Funder Implementation Study to demonstrate how existing technologies can be combined to allow Funders to realistically mandate, and grantees to realistically produce FAIR research outputs.”

Open Research Funders Group

<http://www.orfg.org/>

“Federal grant recipients should, first and foremost, be expected to deposit their data in a data environment that supports the FAIR data sharing principles - findable, accessible, interoperable, and reusable. The FAIR principles are at the core of the open data and reproducibility movement. Any repository housing Federally supported data should clearly and publicly articulate how it conforms to the core components of FAIR”



<https://datascience.nih.gov/data-ecosystem>

“A guiding principle of the NIH Strategic Plan for Data Science, and the approach of the ODSS, is that all biomedical research data should adhere to FAIR principles. This means data should be findable, accessible, interoperable, and reusable (FAIR). While there is no one-size-fits-all solution, ODSS advocates for alignment to the FAIR principles across the biomedical research enterprise.”

FAIR in industry

FAIRplus (<https://fairplus-project.eu/>)

“The FAIRplus project aims to develop tools and guidelines for making life science data FAIR (Findable, Accessible, Interoperable, Reusable). The project has 22 partners from academia and industry”

Goals

- To establish a process for selecting and prioritising IMI project databases for FAIRification.
- To develop guidelines, tools and metrics needed to make data Findable, Accessible, Interoperable and Reusable (FAIR).
- To deliver tailored training for data handlers (academia, SMEs and pharmaceuticals).
- To change and sustain the data management culture in pharma, academia and SMEs.
- To organise FAIR 'Innovation and SME Events' to foster an innovation ecosystem on FAIR open data to power future reuse, knowledge generation, and societal benefit.

Researchers

- Meeting funder requirements
- Make your data discoverable, especially to potential collaborators and others doing the same work
- Create data that adheres to standards and formats, making it interoperable
- Add metadata
 - Related scripts
 - Terms from discipline-related ontologies

Librarians

- Keep up with standards and ontologies
- Metadata librarians work with disciplinary librarians to understand the FAIR-based needs
- Know what the good repositories are

<https://www.coretrustseal.org/>

https://datascience.nih.gov/trusted_digital_repository

ALL OF THIS WILL CHANGE

10 FAIR Data & Software Things

Nanotechnology

Astronomy

Linked Open Data

Imaging

Music

The European Open Science Cloud (EOSC)

Oceanography

Research Software

Research Libraries

10 FAIR Data & Software Things

Research Data Management Support

International Relations

Humanities: Historical Research

Geoscience

Biomedical Data Producers, Stewards, and Funders

Biodiversity

Australian Government Data/Collections

Archaeology

Advice

- Go over the FAIR bullet points and collect a group of people that can help you work with them
- Figure out what communities you want or need to find your data, and incorporate their community standards in describing your data
 - Ontology
 - Distinctive metadata fields
- Get PIDs for your datasets
- Know what data standards are applicable to you and use them (dates, geodata)

Challenges

What kinds of metadata do we have in common, and where do we diverge across disciplines?

FAIR Convergence Symposium

https://docs.google.com/presentation/d/1G8J3B520yV48o06Fsnz_5MMW88dPVA5dHcf2DaO0ohM/edit?usp=sharing

Challenges

FAIR are guidelines, not rules

Different people within the organization understand different aspects of FAIR



Communication Protocols Metadata structure



Ontologies Provenance metadata



Disciplinary metadata
Relationships between files
Number of files

How to support FAIR?

FAIR is a shared responsibility

FAIR needs an ecosystem of services

FAIR needs people

Open (research) needs FAIR



Recommendations

- (1) **funders** and **institutions** should consider FAIR alignment and data sharing as part of research assessment, among other criteria;
- (2) **services** should support domain-specific ontologies by identifying disciplines that lack ontologies and enriching existing registries of ontologies;
- (3) **repositories** should support FAIR data by developing tools, such as APIs, sharing best practices, and undergoing FAIR-aligned certification; and
- (4) **institutions** should support FAIR awareness and implementation by establishing data stewardship programs providing simple and intuitive training for researchers.

Resources

Top 10 FAIR Data & Software Things (<https://librarycarpentry.org/Top-10-FAIR/>)

FAIRSFAIR (<https://www.fairsfair.eu/>)

GO FAIR Europe (<https://www.go-fair.org/go-fair-initiative/>)

Next webinar

GO FAIR US

Thursday, December 10, 2020 – 12-1PM Eastern Time / 9-10AM Pacific Time (4PM UTC)

[Click here to register for the webinar](#)

Additional topics on the horizon:

- FAIR & AI
- FAIR Educational/Training Material
- FAIR Implementation in the Geosciences
- FAIR Implementation Profiles & Data Stewardship Plans/DMPs as Nanopublications
- 3-point FAIRification strategy