

FAIR for research software

Daniel S. Katz (d.katz@ieee.org, @danielskatz)

Chief Scientist, NCSA

Associate Research Professor, CS, ECE, iSchool

University of Illinois at Urbana Champaign

On behalf of the FORCE11/RDA/ReSA FAIR4RS WG – co-leaders: Morane Gruenpeter, Carlos Martinez, Neil Chue Hong, Paula A. Martinez, Michelle Barker, Leyla Jael Castro, Jennifer Harrow, Fotis Psomopoulos

RDA FDO Fabric – Project Sharing

18 August 2022

I ILLINOIS

NCSA | National Center for
Supercomputing Applications

<https://doi.org/10.5281/zenodo.7006955>



The FAIR Principles

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons 

Scientific Data 3, Article number: 160018 (2016) | [Cite this article](#)

194k Accesses | 2450 Citations | 1852 Altmetric | [Metrics](#)

A set of principles, to ensure that data are shared in a way that enables and enhances reuse by humans and machines

Findable

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

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.

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.

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.

FAIR for non-data objects: some context

- FAIR Principles, at a high level, are intended to apply to all research objects; both those used in research and those that are research outputs
- Text in principles often includes "(Meta)data ..."
 - Shorthand for "metadata and data ..."
- Principles applied via dataset creators and repositories, collectively responsible for creating, annotating, indexing, preserving, sharing the datasets and their metadata
 - Assumes separate and sequential creator/publisher (repository) roles
- What about non-data objects?
 - While they can often be stored as data, they are not just data
- While high level goals (F, A, I, R) are mostly the same, the details and how they are implemented depend on
 - How objects are created and used
 - How/where the objects are stored and shared
 - How/where metadata is stored and indexed
- Work needed to define, then implement, then adopt principles

Need for FAIR for non-data objects

- FAIR Principles, are intended to apply to all digital objects (Wilkinson et al. 2016)

Recommendation 5:

*Recognise that FAIR guidelines will require **translation for other digital objects** and support such efforts.*

2020: ‘Six Recommendations for Implementation of FAIR Practice’
(FAIR Practice Task Force EOSC, 2020)

Recommendation 2:

*Make sure **the specific nature of software** is recognized and not considered as “just data” particularly in the context of discussion about the notion of FAIR data.*

2019: Opportunity Note by French national Committee for Open Science's Free Software and Open Source Project Group
(Clément-Fontaine, 2019)

- We focused on **adaptation and adoption** of the FAIR principles to research software

FAIR for non-data objects: some efforts

FAIR Principles for Research Software (FAIR4RS Principles)

16
Mar
2022

By Neil Chue Hong

FAIR for Research Software (FAIR4RS) WG



January 01 2020

FAIR Computational Workflows

Carole Goble, Sarah Cohen-Boulakia, Stian Soiland-Reyes, Daniel Garijo, Yolanda Gil, Michael R. Crusoe, Kristian Peters, Daniel Schober

> Author and Article Information

Data Intelligence (2020) 2 (1-2): 108–121.

https://doi.org/10.1162/dint_a_00033



Breakout 7 Data Infrastructures - Organisa... The FAIR Agenda WGs Getting started

WG FAIR for Virtual Research Environments: FAIR for VREs - The Path Forward

7:30 AM - 9:00 AM

Room E

Ten simple rules for making training materials FAIR

Leyla Garcia, Bérénice Batut, Melissa L. Burke, Mateusz Kuzak, Fotis Psomopoulos, Ricardo Arcila, Teresa K. Attwood, Niall Beard, Denise Carvalho-Silva, Alexandros C. Dimopoulos, Victoria Dominguez del Angel, Michel Dumontier, Kim T. Gurwitz, [...], Patricia M. Palagi [view all]

Published: May 21, 2020 • <https://doi.org/10.1371/journal.pcbi.1007854>



Steps towards defining FAIR principles for Machine Learning (ML)

Home

28
JUL
2021

Steps towards defining FAIR principles for Machine Learning (ML)

Submitted by Fotis Psomopoulos



Software vs. data

- Software is data, but it's not just data
 - Software is executable, data is not
 - Data provides evidence, software provides a tool
 - Software is a creative work, scientific data are facts or observations
 - Different licensing and copyright practices
 - Software suffers from a different type of bit rot (collapse) than data
 - The lifetime of software is generally not as long as that of data
 - For open source, no natural sequential creator/publisher process & no natural publisher (repository)

D. S. Katz et al., "Software vs. data in the context of citation," PeerJ Preprints 4:e2630v1, 2016. <https://doi.org/10.7287/peerj.preprints.2630v1>

FAIR for Research Software (FAIR4RS)

- Working group defining FAIR principles for research software
 - Led by Michelle Barker, Neil Chue Hong, Leyla Garcia, Morane Gruenpeter, Jennifer Harrow, Daniel S. Katz, Carlos Martinez, Paula A. Martinez, Fotis Psomopoulos
 - Additional subgroup leaders: Tom Honeyman, Anna-Lena Lamprecht, Chris Erdman, Sandra Gesing, Qian Zhang



Defining research software

- **Research Software** includes source code files, algorithms, scripts, computational workflows and executables that were **created during the research process or for a research purpose**
- Additional software components (e.g., operating systems, libraries, dependencies, packages, scripts, etc.) that are used for research but were **not created during or with a clear research intent should be considered software in research** and not Research Software
- This differentiation may vary between disciplines

<https://doi.org/10.5281/zenodo.5504016>

FAIR₄RS Principles

Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A-L, Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., Honeyman, T., et al. (2022). FAIR Principles for Research Software version 1.0. (FAIR₄RS Principles v1.0). Research Data Alliance. DOI: [10.15497/RDA00068](https://doi.org/10.15497/RDA00068)

- Findable: Software, and its associated metadata, is easy for both humans and machines to find.
- Accessible: Software, and its metadata, is retrievable via standardized protocols.
- *Interoperable: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.*
- *Reusable: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).*

(key differences from FAIR data principles in *italics*)

Findable

Findable: Software, and its associated metadata, is easy for both humans and machines to find.

F1. Software is assigned a globally unique and persistent identifier.

F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.

F1.2. Different versions of the software are assigned distinct identifiers.

F2. Software is described with rich metadata.

F3. Metadata clearly and explicitly include the identifier of the software they describe.

F4. Metadata are FAIR, searchable and indexable.

Accessible

Accessible: Software, and its metadata, is retrievable via standardized protocols.

A1. Software is retrievable by its 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 software is no longer available.

Interoperable

Interoperable: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.

I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.

I2. Software includes qualified references to other objects.

Reusable

Reusable: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).

R1. Software is described with a plurality of accurate and relevant attributes.

R1.1. Software is given a clear and accessible license.

R1.2. Software is associated with detailed provenance.

R2. Software includes qualified references to other software.

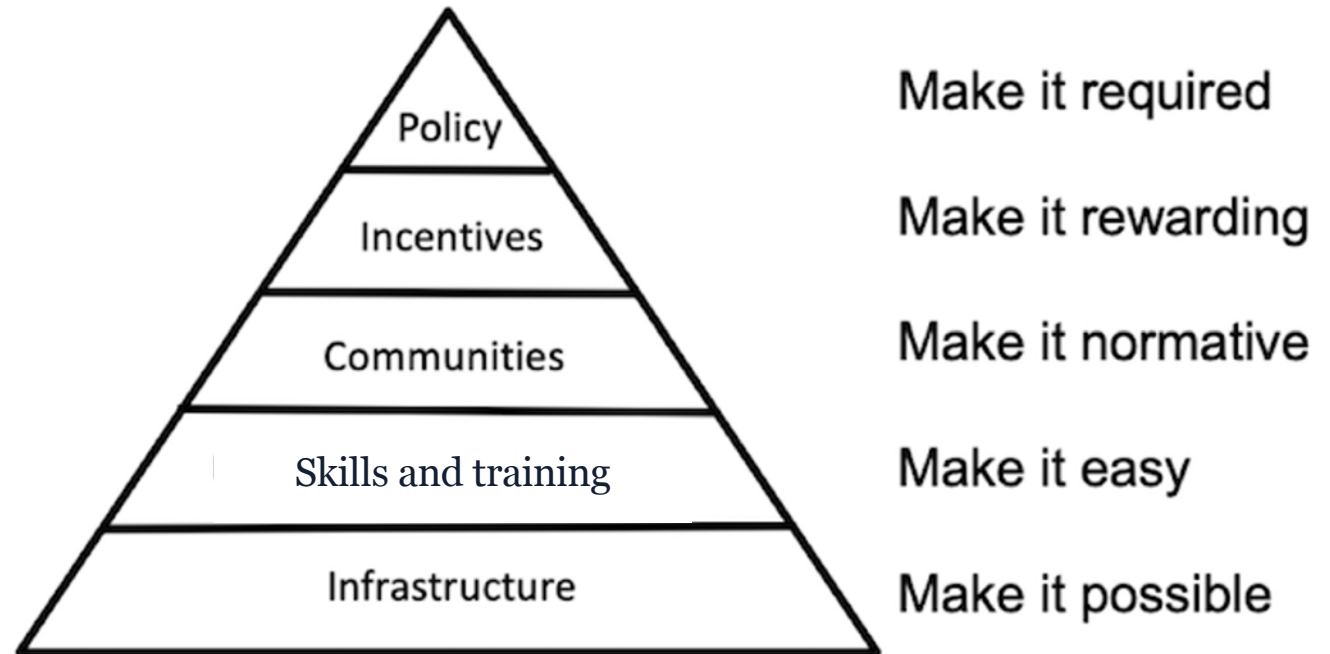
R3. Software meets domain-relevant community standards.

Who is expected to apply FAIR4RS?

And why?

“...the application of the FAIR4RS Principles is the responsibility of the owners (who are often the creators) of the software, not the users.”

“The FAIR4RS Principles must also be enabled and supported by various stakeholders in the larger ecosystem that supports research software (e.g., repositories and registries).”



*Adapted by Michelle Barker from original by Brian Nosek:
Strategy for Culture Change (2019)*

Using FAIR4RS and what's next

- Survey of adoption guidelines: <https://doi.org/10.5281/zenodo.6374598>
- Study of adopting organizations: <https://doi.org/10.5281/zenodo.6258366>
- Desired stakeholder actions:
 - Scholarly societies and librarians: develop guidance aimed at their communities
 - Individuals and software projects: make their software FAIR
 - Publishers: require FAIR software
 - Funders: require FAIR software
 - Institutions: incentivize and evaluate their employees based, in part, on the FAIRness of the software they produce

FAIR4RS governance

- Governance (interpretation, future revisions) turned over to RDA Software Source Code Interest Group
 - Concerns/queries about the principles can be raised at SSC IG events at RDA plenaries, where adopters can report back on progress
 - Plan is to review of principles in 2 years, and possibly update (if needed)
 - Full maintenance & retirement plan for the principles on RDA website

Future work beyond FAIR4RS principles

- FAIR4RS exposes ecosystem gaps, particularly related to metadata, archiving, versions
 - Creator/publisher sequence doesn't typically apply
 - Where is metadata stored? (in code repository for open source?, for closed source?, in archival repository?, in registry?)
 - Where is code archived? (GitHub/Gitlab are not archival, registries are not archival, repositories? Software Heritage?)
 - Different use cases need specific version, latest version, all versions
- Lots of work beyond FAIR: quality, correctness, reproducibility, openness, ...

FAIR DO and research software

- My potential concerns
 - A failing of FAIR was its bias towards data and just data, at a detailed level
 - Now we have lots of different groups creating new FAIR principles for different types of objects, which eventually I think we will need to unify
 - FAIR DO seems to be repeating this
 - “The International FDO Forum will bring together infrastructure researchers, engineers and developers ..., digital research artifacts, information technology researchers, tool builders, and data/service providers ..., researchers and data users ..., policy makers, funders, and leading thinkers ...”
 - Where are software and software developers?
 - FAIR DO seems to be a misleading name
 - “We aim toward a new infrastructure, an architectural extension of the Internet – the Internet of FAIR Data and Services”
 - What are these digital objects but again data?
 - What “services”?
 - How does “an architectural extension of the Internet” fit in here?
- (quotes from <https://fairdo.org/about/>)
- But maybe I’m wrong
- In either case, members of the FAIR4RS community want to help – how can we?

Acknowledgements

We thank all contributors, working on this since VP16!

- All 261 members and ~233 participants of the FAIR for Research Software Working group #FAIR4RS
- Present and past steering committee members for coordinating a range of activities
- Alfred P. Sloan Foundation and Wellcome Trust for support

<https://www.rd-alliance.org/group/fair-4-research-software-fair4rs-wg>

The FAIR4RS WG is now complete!