

FAIR 4 Research Software (FAIR4RS)

RDA France

Morane Gruenpeter (Software Heritage, Inria - FAIRsFAIR)

Carlos Martinez, Neil Chue Hong, Daniel S. Katz, Paula A. Martinez, Michelle Barker,
Leyla Jael Castro, Jennifer Harrow, Fotis Psomopoulos

research data sharing without barriers
rd-alliance.org

15th October, 2021

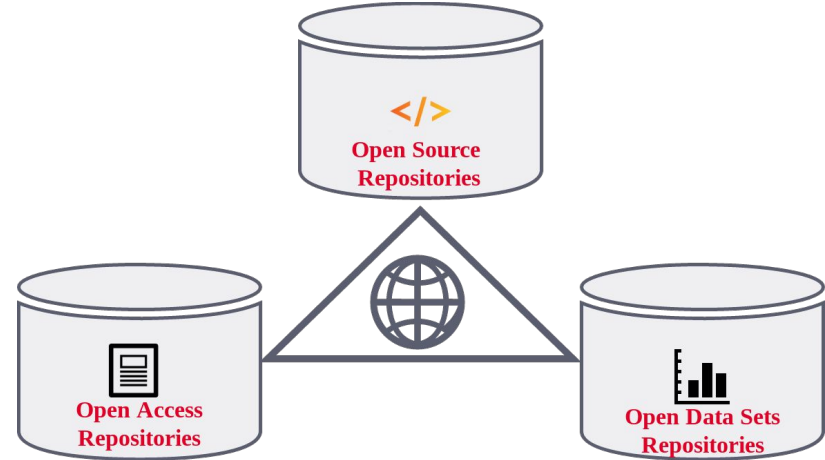
Today's goals

- Introduction
 - Research software
 - Present the FAIR4RS WG activities
- Review FAIR principles for research software
- Next steps

Software in Research: A pillar of Open Science

Multiple facets, it can be seen as:

- a **tool**
- a research **outcome** or result
- **the object** of research

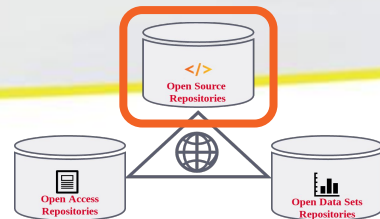


*Three pillars of Open Science, Software Heritage
CC-By 4.0 2019*



Software Heritage

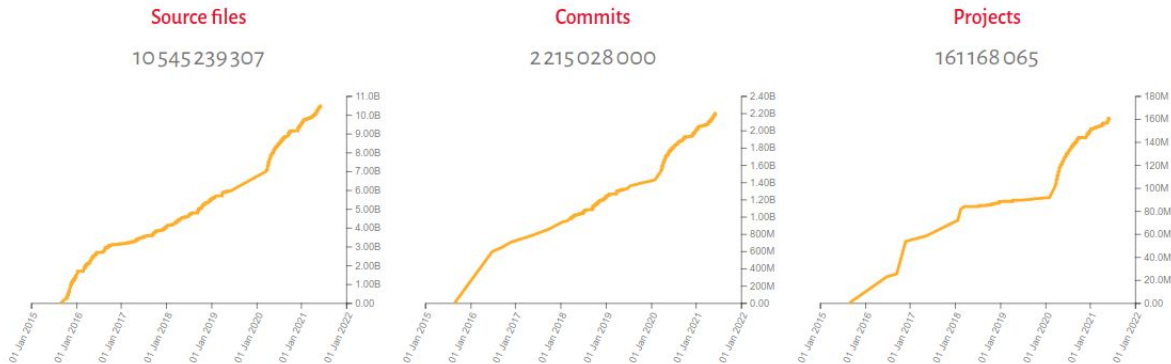
THE GREAT LIBRARY OF SOURCE CODE



Three pillars of Open Science,
Software Heritage CC-BY 4.0 2019

Collect, preserve and share all software source code

Preserving our heritage, enabling better software and better science for all



Source: Software Heritage
(June 2021)

Why are we here? A plurality of needs

Researchers

- **archive and reference** software used and created in articles
- **find** useful software
- **get credit** for developed software
- **verify/reproduce/improve** results

Laboratories/teams

- **track** software contributions
- **produce** reports
- **maintain** web page

Research Organization

know its **software assets** for:

- technology transfer,
- impact metrics,
- strategy

Open science/source in France

L'ouverture
des codes sources
des logiciels est
un enjeu majeur
de reproductibilité
des résultats
scientifiques

Deuxième Plan national pour la science ouverte



Définir et promouvoir une politique en matière de logiciels libres

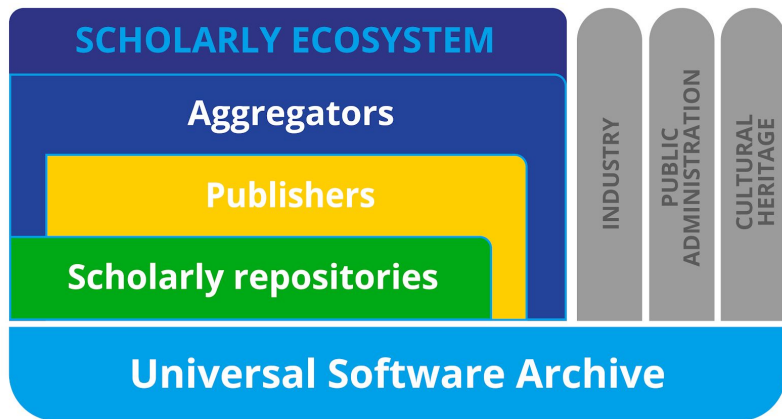
- Établir une **Charte nationale des logiciels libres** issus de l'enseignement supérieur, de la recherche et de l'innovation.
- Développer le lien entre données et logiciels grâce au réseau des **administrateurs des données, des algorithmes et des codes sources** dans les établissements.
- Émettre des **recommandations auprès des organismes financeurs** pour accompagner au mieux le développement logiciel.
- Faire monter en compétence les structures de valorisation sur **les modèles économiques associés** à la production de logiciels libres.
- Soutenir **Software Heritage** et recommander son adoption pour l'archivage et le référencement des codes sources.

Reconnaître les codes sources comme une contribution à la recherche

- Créer un **prix du logiciel libre pour la recherche** qui récompense les équipes et projets exemplaires dans le domaine.
- **Mieux valoriser** les productions logicielles dans la carrière des chercheurs, des personnels d'accompagnement à la recherche et dans l'évaluation des structures de recherche.
- **Suivre dans le temps** la production de codes et logiciels de la recherche française pour en identifier les dynamiques, l'ouverture et les impacts grâce au baromètre de la science ouverte.
- Construire un **catalogue** des logiciels issus de la recherche en utilisant un schéma de métadonnées normalisé et partagé entre tous les acteurs de l'enseignement supérieur, de la recherche et de l'innovation.

Source: [Ministère de l'enseignement supérieur et de la recherche](https://www.enseignementsup-recherche.gouv.fr/fr/le-deuxieme-plan-national-pour-la-science-ouverte)

EOSC Scholarly Infrastructures for RS (SIRS)



Four pillars: Archive, Reference, Describe, Credit

EOSC Scholarly Infrastructures for Research Software

- Chairs
 - Roberto Di Cosmo, Software Heritage, Inria
 - José Benito Gonzalez Lopez, Zenodo, CERN

Important policy tool in Open Science

- 9 infrastructures
 - 3 archives
 - 3 open access publishers
 - 3 aggregators
- recommendations
 - archive in Software Heritage,
 - use SWHID
 - open non profit
 - default to open source for research software

"all research software should be made available under an Open Source license by default, and all deviations from this default practice should be properly motivated"

See <https://doi.org/10.2777/28598>

What is at stake? In order of difficulty

Archive

- Research software artifacts must be properly archived
- make it sure we can retrieve them (reproducibility)

Reference

- Research software artifacts must be properly referenced
- make it sure we can identify them (reproducibility)

Describe

- Research software artifacts must be properly described
- make it easy to discover them (visibility)

Cite (for credit)

- Research software artifacts must be properly cited (not the same as referenced!)
- to give credit to authors (evaluation!)

Software in RDA

Birds of a Feathers RDA P9, [Barcelona, April 2017](#)

- Software Source code: Sharing, Preservation and Reproducibility
- 60 participants

Software groups:

- RDA, ReSA and FORCE11 [FAIR for Research Software WG](#) | 3.11 | 14:00 - 15:30 UTC | **Breakout 2** [session](#)
- RDA [Curating for FAIR and Reproducible Research](#) (CURE FAIR WG) | 3.11 | 22:30 - 00:00 UTC | **Breakout 3** [session](#)
- RDA [Software Source Code IG](#) (SSC IG) | 9.11 | 16:00 - 17:30 UTC | **Breakout 8** [Source code and reproducibility](#)

3.11 | 22:30 - 00:00 | **Breakout 3** *BoF - Skills and training curriculums to support FAIR for Research Software*

11.11 | 15:00 - 16:00 UTC | **Plenary session: Data and Software Sharing: The Role of Societies and Associations**

Software is not just another type of data

Recommendation n°5 :

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

2019: 'Six Recommendations for Implementation of FAIR Practice'

([FAIR Practice TF, 2020](#))

Recommendation n° 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: the Opportunity Note by the French national Committee for Open Science's Free Software and Open Source Project Group
([Clément-Fontaine, 2019](#))

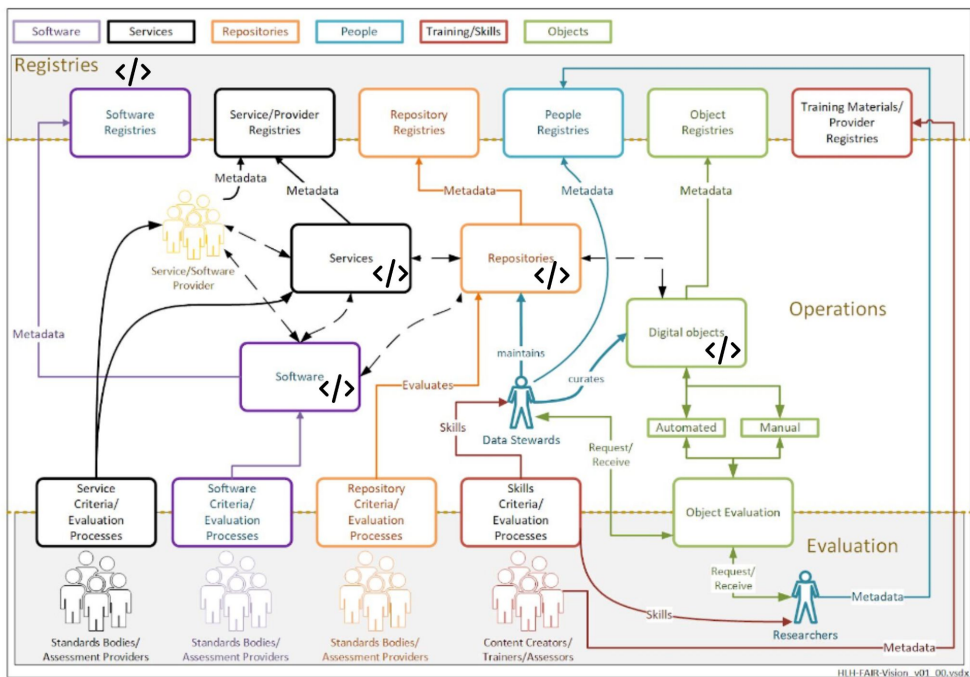
Software in the FAIR ecosystem

[10.5281/zenodo.4095092](https://doi.org/10.5281/zenodo.4095092)

Webinar FAIR + Software: decoding the principles (Nov 2020)

October 16, 2020

Project milestone Open Access



M2.15 Assessment report on 'FAIRness of software'

Grünepeter, Morane; Di Cosmo, Roberto; Koers, Hylke; Herterich, Patricia; Hooft, Rob; Parland-von Essen, Jessica; Tana, Jonas; Aalto, Tero; Jones, Sarah

Software has an important place in academia and as such it has an important place in the FAIR ecosystem. Software can be used throughout the research process; however it can also be an outcome of the research process. Distinguishing between these different roles is essential for any assessment of the 'FAIRness of software'.

This is the first milestone of the FAIRsFAIR project focused specifically on software as a digital object. In this report we discuss the state-of-the-art of software in the scholarly ecosystem in general and in the FAIR literature in particular. We identify the challenges of different stakeholders when it comes to finding and reusing software. Furthermore, we present an analysis of nine resources that call for the recognition of software in academia and that present guidelines or recommendations to improve its status - either by becoming more FAIR or by improving the curation of software in general. With this analysis we demonstrate to what extent each of the FAIR principles is seen as relevant, achievable and measurable, and in what sense it benefits software artifacts. Finally, we present 10 high-level recommendations for organizations that seek to define FAIR principles or other requirements for research software in the scholarly domain.

Feedback and suggestions will be most welcome as comments on the public Google Doc version of this report <https://docs.google.com/document/d/1yvdLP6oH3XozV4CJtThzGNHkscBDvmxfuDLB6Q/edit?usp=sharing>



Ecosystem components, to highlight the software roles in the Ecosystem, the symbol </> was added (Original diagram 3 from L'Hours & Von Stein, 2020)

research data sharing without barriers
rd-alliance.org

FOR
The Future of Research Communication

Are services considered software?

Technically-oriented

FAIR
enablement

Quality of
service

Open &
Connected

Socially-oriented

User centrality

Transparency

Longevity

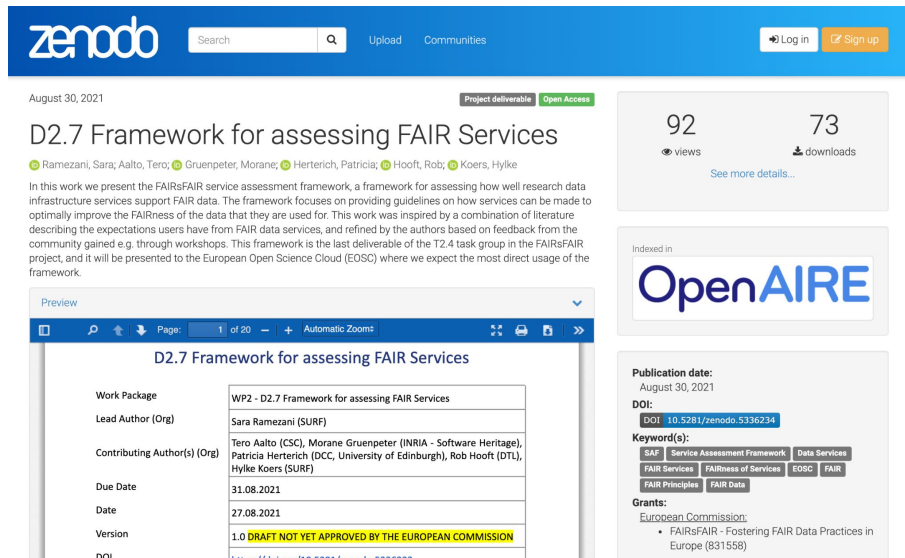
Ethical & Legal

FAIR enablement:

Enable / Respect / Reduce

Enable: Augment / Facilitate

FAIRsFAIR - WP2 - Task 2.4



The screenshot shows the Zenodo project page for 'D2.7 Framework for assessing FAIR Services'. The page includes a header with the Zenodo logo, search bar, and navigation links. The main content area displays the project title, authors (Ramezani, Sara; Aalto, Tero; Gruenpeter, Morane; Herterich, Patricia; Hooft, Rob; Koers, Hylke), and a description of the framework. A table lists project details: Work Package (WP2 - D2.7 Framework for assessing FAIR Services), Lead Author (Sara Ramezani), Contributing Author(s) (Tero Aalto, Morane Gruenpeter, Patricia Herterich, Rob Hooft, Hylke Koers), Due Date (31.08.2021), Date (27.08.2021), Version (1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION), and DOI (https://doi.org/10.5281/zenodo.5336234). The page also shows 92 views and 73 downloads, and a link to the OpenAIRE record.

D2.7 Framework for assessing FAIR Services

<https://doi.org/10.5281/zenodo.5336234>

FAIR for Research Software (FAIR4RS) WG

Main objective

Defining **FAIR principles for research software**

Steering committee and WG chairs:

Morane Gruenpeter, Paula A. Martinez, Carlos Martinez, Michelle Barker, Daniel S. Katz, Leyla Garcia, Neil Chue Hong, Fotis Psomopoulos and Jennifer Harrow

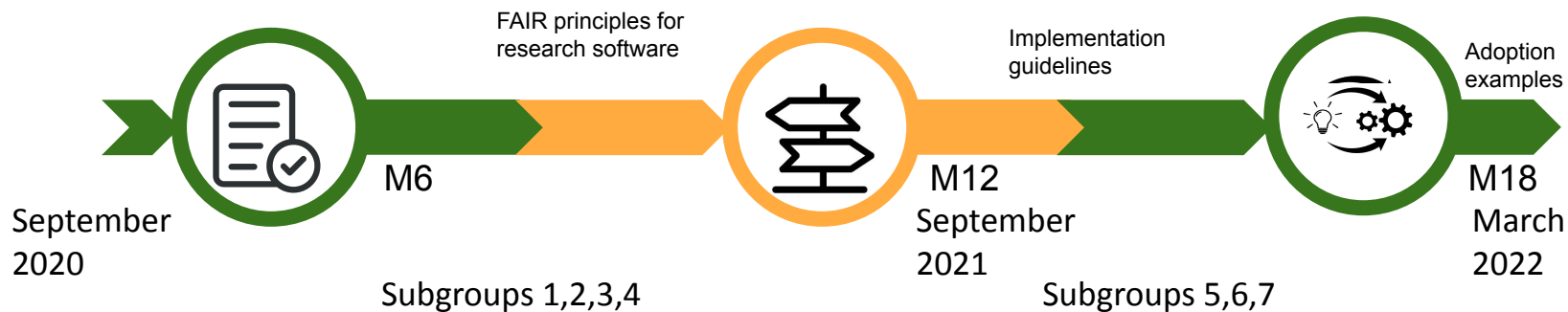


- Acknowledging the ~228 members and contributors of the FAIR for Research Software working group #FAIR4RS

[Join the WG](#)

#FAIR4RS timeline

Endorsed by RDA
Sep 2020



FAIR4RS initial subgroup activities and outputs

- 1. [A fresh look at FAIR for Research Software](#) examined the FAIR principles in the context of research software from scratch, not based on pre-existing work. **Lead: Daniel S. Katz**
- 2. [FAIR work in other contexts](#) examined efforts to apply FAIR principles to different forms including workflows, notebooks and training material, to provide insights for the definition and implementation of FAIR principles for research software. **Lead: Michelle Barker**
- 3. [Defining Research Software: a controversial discussion](#) reviews existing definitions of research software in order to provide the overall context of the subgroup outputs. **Lead: Morane Gruenpeter**
- 4. [Review of new research related to FAIR Software](#) reviewed new research around FAIR software that has come out since the release of the [Towards FAIR principles for research software](#) (Lamprecht et al., 2019). **Lead: Neil Chue Hong**

Defining the scope

Research Software includes source code files, algorithms, scripts, computational workflows and executables that were created during the research process or for a research purpose. 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. The minimal requirement for achieving computational reproducibility is that all the computational components (Research Software, software used in research, documentation and hardware) used during the research are identified, described, and made accessible to the extent that is possible.

Research Software \neq Software in research

First FAIR4RS community consultation

<http://doi.org/10.5281/zenodo.4635410>

First community consultation to get feedback on findings of subgroups, and questions around scope of the draft FAIR4RS principles

- **24 February - 10 March 2021**
- 215 comments from 19 named contributors (other than the SC) + other anonymous contributors.

Used as main input for FAIR4RS drafting sprints, with questions and clarifications raised by the community discussed by the drafting team to determining the intent of the principles

March 25, 2021

FAIR4RS WG subgroup community consultation March 2021

👤 Katz, Daniel S.; 👤 Chue Hong, Neil P.; 👤 Barker, Michelle; 👤 Gruenpeter, Morane

Distributor(s)

👤 Martinez, Paula Andrea

Project member(s)

👤 Psomopoulos, Fotis; 👤 Garcia Castro, Leyla Jael; 👤 Martinez-Ortiz, Carlos; 👤 Harrow, Jennifer; 👤 Tovani-Palone, Marcos Roberto; 👤 Honeyman, Tom; 👤 Leng, Joanna; 👤 Sinha, Manodeep; 👤 Islam, Sharif; 👤 Loewe, Axel; 👤 Fox, Merc; 👤 Atmojo, Udayanto Dwi; 👤 Pollard, Tom; 👤 McNally, James; 👤 Sandström, Malin; 👤 van Werkhoven, Ben; 👤 Todorov, Ilian; 👤 Herterich, Patricia; 👤 Shanahan, Hugh; 👤 Servillat, Mathieu; 👤 Rangelova, Elena; 👤 Jones, Catherine; 👤 Tartarini, Daniele; FAIR4RS WG

This document is the result of the four subgroups of the FAIR for Research Software working group, which is working under the [Research Data Alliance](#), the [Research Software Alliance](#), and [FORCE11](#). These subgroups independently examined the FAIR principles in relation to software.

FAIR4RS-subgroup1 started with the original FAIR principles (Wilkinson et al. 2016) and worked to

1. determine what part of the original FAIR principles apply as is to research software;
2. determine what part of the original FAIR principles doesn't apply at all to research software; and
3. determine what part of the original FAIR principles applies to research software, but with a different definition or different details, starting with the original FAIR principles themselves, and not relying on work done by others to apply them to research software, such as by [Lamprecht et al. \(2020\)](#).

This led to [\(the preprint Katz et al. 2021\)](#) and the opinion paper [Katz et al. 2021](#) that includes:

- a discussion of the differences between software and data,
- an initial straightforward translation that was collected from the FAIR4RS-subgroup1 participants;
- a discussion about the nuances of the currently defined rules in the context of research software;
- a proposed set of principles adapted to the FAIR research software case;
- a comparison of those proposed principles with the FAIR data principles;
- a set of gaps in our current infrastructure and existing practices that make implementing the proposed principles difficult;

Other Open Access

98

👁 views

67

📄 downloads

[See more details...](#)

Indexed in

OpenAIRE

Publication date:

March 25, 2021

DOI:

DOI: [10.5281/zenodo.4635410](https://doi.org/10.5281/zenodo.4635410)

Keyword(s):

software research software FAIR Findable
Accessible Interoperable Reusable
Community consultation FAIR4RS

Grants:

[Research Councils UK:](#)

- The Software Sustainability Institute: Phase 2 (EP/N006410/1)
- The UK Software Sustainability Institute: Phase 3 (EP/S021779/1)

Full community review


<https://doi.org/10.15497/RDA00065>

Draft for formal community review of the draft FAIR4RS principles

- **Review period:** Friday, 11 June, 2021 to Sunday, 11 July, 2021


During this period the WG actively engaged with the community to respond to their feedback and capture their comments in the best possible way.

Drafting group meet in August to address comments received.



Author: Limor Peer
Date: 15 Jun, 2021

Thank you for producing a very comprehensive and clear document. I'm pleased to see language in this version that refers to the shared responsibility for applying FAIR4RS Principles -- I think it's important to emphasize that while the primary responsibility lies with software creators and owners, it often falls to those tasked with quality review and stewardship (who are really the first users) to follow through. I suggest also referencing this issue, and the need to build capacity for this type of work, in the section on the path to adoption. Thanks again for great work!



Author: Neil Chue Hong
Date: 25 Jun, 2021

Dear Limor,
on behalf of the FAIR4RS drafting team, thank you for your comment.
You make a very good point, and we will add a reference to this issue in the section on path to adoption as you suggest.



Author: Yo Yehudi
Date: 21 Jun, 2021

These principles are very clear and well laid out - two small comments, both about possible examples of the principles:

A1 talks about protocols to access software. I wasn't sure if this meant something like git or https, or whether it meant a defined process document on a website, or something else. Maybe it meant all of those? :)

Similarly, F4: Metadata are FAIR and indexable. I couldn't decide based on this if publishing a software artifact as a ZIP on zenodo, with embedded .cff might be compliant with this rule, or if perhaps I am supposed to upload the cff itself to a repo somewhere... or maybe something else? I broadly understand the _intention_ of this rule but struggled a little to understand the specifics about how one might meaningfully comply.

Other than that I thought the rules were really clear and was delighted to see the note about overloading accessibility as a term :) It's too little loved as it is and I dread seeing it dropped in favour of FAIR accessibility.



Author: Neil Chue Hong
Date: 25 Jun, 2021

Community engagement

60+ events

July 2020

Dec 2021

Locations:

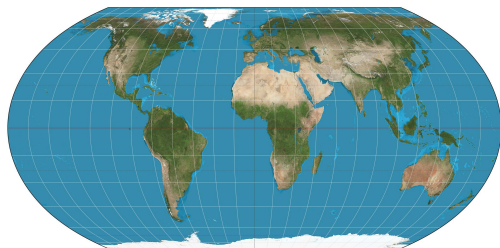
US, UK, NZ, CA, NL, FR,
MX, “Africa”, “Nordics”,
and more!

Channels:

Mailing list, Webinars,
Twitter, RDA website,
GitHub, etc.

Events:

CarpentryCon, SORSE,
IEEE eScience, CW21,
Collegeville, Utrecht
University, PIDapalooza,
INCF, etc.



<https://github.com/force11/FAIR4RS/blob/master/CommunityEngagement.md>

Development of the FAIR4RS Principles

- Intent and methods of the FAIR Guiding Principles taken as starting point:
 - “maximize the added-value gained by contemporary, formal scholarly digital publishing”
 - “to ensure transparency, reproducibility, and reusability.”
- The FAIR Principles are aspirational, and FAIR is not binary
 - The aim of FAIR (and FAIR) metrics is to show progress to **increasing FAIRness**
- Software encompasses many forms, which may benefit different users
 - **Source code** is often the most useful form to understand the software, and the easiest form to apply the FAIR4RS Principles.
- Many software engineering practices are relevant to the FAIR4RS Principles
 - For instance: localization can improve findability, design patterns can improve interoperability, and documentation and encapsulation can improve reusability.
 - Nevertheless, while important more generally for producing high quality software, they are best addressed separately from (but as a complement to) the FAIR4RS Principles.

FAIR Principles for Research Software

[10.15497/RDA00065](https://doi.org/10.15497/RDA00065)
[FAIR4RS RDA webinar](#)

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

F1. Software is assigned a globally unique and persistent identifier

- F1.1. Different components of the software are assigned distinct identifiers representing different levels of granularity
- F1.2. Different versions of the same 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 and are searchable and indexable

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: Software interoperates with other software through 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: Software is both usable (it can be executed) and reusable (it 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

FAIR4RS WG. (2021, June). FAIR Principles for Research Software



FORCE11

The Future of Research Communications and e-Scholarship

<ReSA>

Research Software Alliance



RESEARCH DATA ALLIANCE

| FAIR Guiding Principles (2016) | FAIR4RS Principles (2021) |
|---|---|
| F. Findable | |
| The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process. | Software, and its associated metadata, is easy to find for both humans and machines. |
| F1. (Meta)data are assigned a globally unique and persistent identifier | F1. Software is assigned a globally unique and persistent identifier. |
| | F1.1. Different components of the software are assigned distinct identifiers representing different levels of granularity. |
| | F1.2. Different versions of the same software are assigned distinct identifiers. |
| F2. Data are described with rich metadata (defined by R1 below) | F2. Software is described with rich metadata. |
| F3. Metadata clearly and explicitly include the identifier of the data they describe | F3. Metadata clearly and explicitly include the identifier of the software they describe. |
| F4. (Meta)data are registered or indexed in a searchable resource | F4. Metadata are FAIR and is searchable and indexable. |

| FAIR Guiding Principles (2016) | FAIR4RS Principles (2021) |
|---|---|
| A. Accessible | |
| Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation. | Software, and its metadata, is retrievable via standardized protocols. |
| A1. (Meta)data are retrievable by their identifier using a standardized communications protocol | A1. Software is retrievable by its identifier using a standardized communications protocol. |
| A1.1. The protocol is open, free, and universally implementable | A1.1. The protocol is open, free, and universally implementable. |
| A1.2. The protocol allows for an authentication and authorization procedure, where necessary | 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 | A2. Metadata are accessible, even when the software is no longer available. |

Interoperable

[10.15497/RDA00065](https://doi.org/10.15497/RDA00065)

| FAIR Guiding Principles (2016) | FAIR4RS Principles (2021) |
|---|---|
| I. Interoperable | |
| The data usually needs to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing. | Software interoperates with other software through exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards. |
| I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. | I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards. |
| I2. (Meta)data use vocabularies that follow FAIR principles | |
| I3. (Meta)data include qualified references to other (meta)data | I2. Software includes qualified references to other objects. |

| FAIR Guiding Principles (2016) | FAIR4RS Principles (2021) |
|--|--|
| R. Reusable | |
| The ultimate goal of FAIR is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings. | Software is both <u>usable</u> (it can be executed) and <u>reusable</u> (it can be understood, modified, built upon, or incorporated into other software). |
| R1. (Meta)data are richly described with a plurality of accurate and relevant attributes | R1. Software is described with a plurality of accurate and relevant attributes. |
| R1.1. (Meta)data are released with a clear and accessible data usage license | R1.1. Software is given a clear and accessible license. |
| R1.2. (Meta)data are associated with detailed provenance | R1.2. Software is associated with detailed provenance. |
| R1.3. (Meta)data meet domain-relevant community standards | R3. Software meets domain-relevant community standards. |
| | R2. Software includes <u>qualified references</u> to other <u>software</u>. |

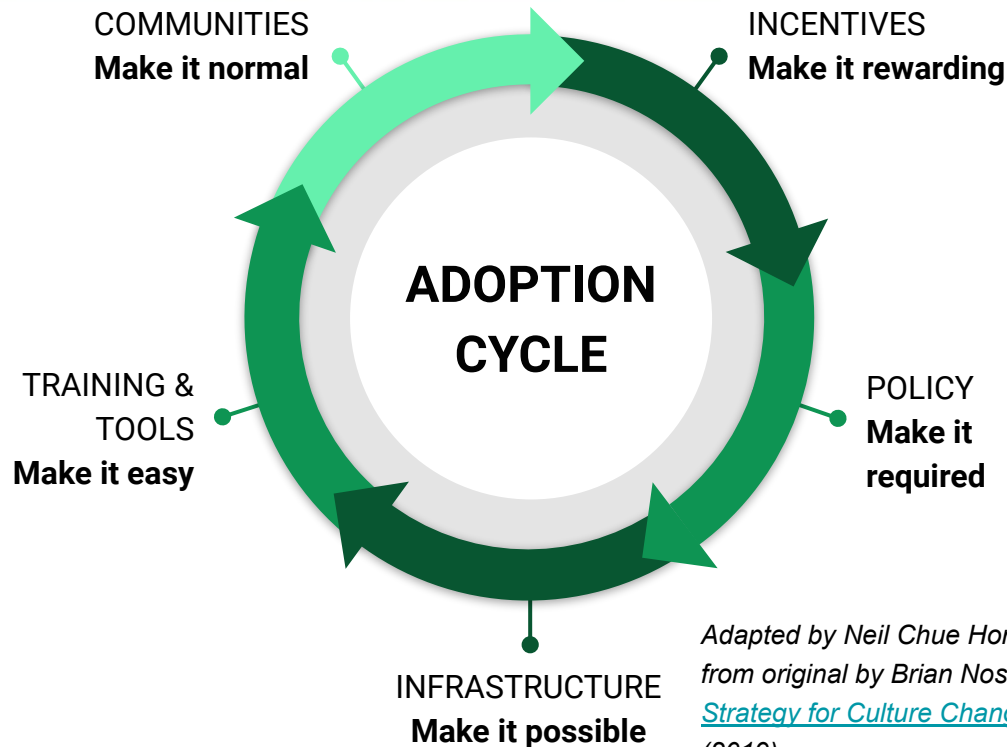
Who is responsible for FAIR software?

Who is expected to apply FAIR?

- 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.”

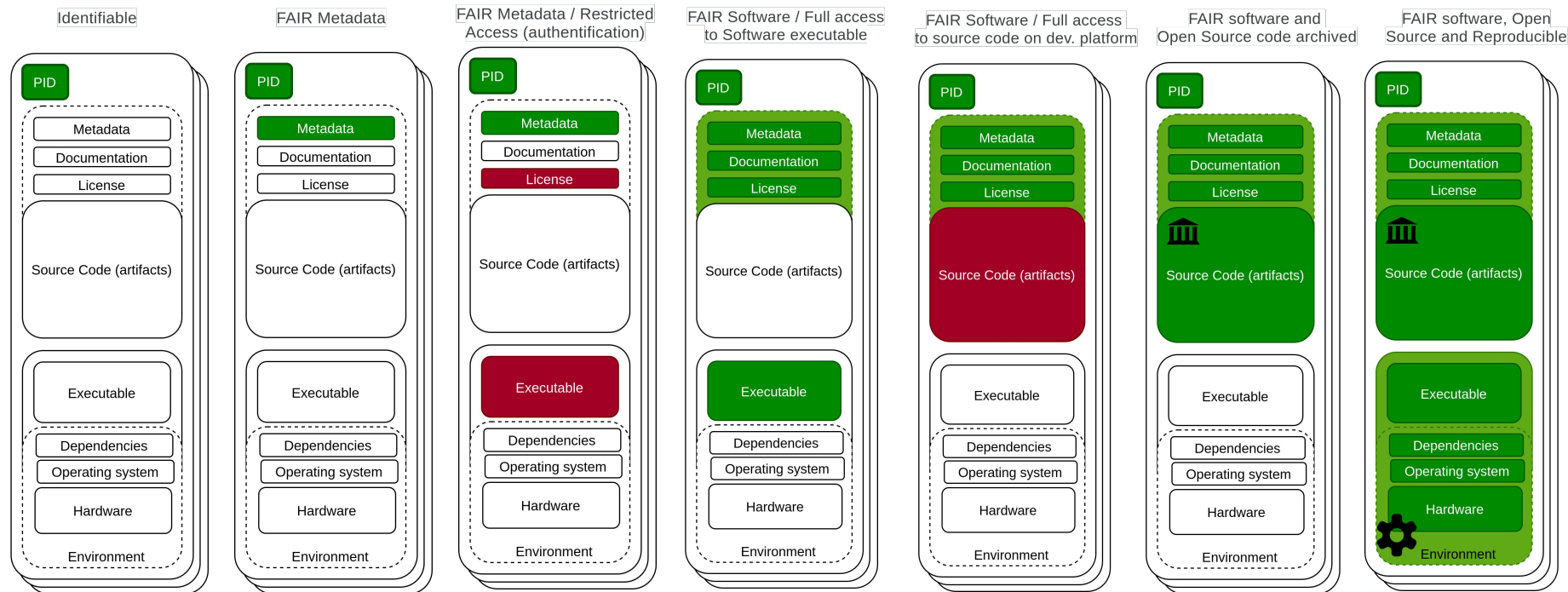
“The FAIR4RS Principles are also relevant to the larger ecosystem and various stakeholders that support research software (e.g., repositories and registries).”



Adapted by Neil Chue Hong
from original by Brian Nosek:
[Strategy for Culture Change](#)
(2019)

Beyond FAIR: FAIR is not the end goal

[10.15497/RDA00065](https://doi.org/10.15497/RDA00065)



(Katz et al. 2020) arXiv:2101.10883

New subgroups

Three new subgroups now starting

- 5. Adoption guidelines
 - Identify, create, review existing resources that facilitate the adoption of FAIR4RS principles
- 6. Adoption support
 - Identify & start to work with organisations following FAIR4RS guidelines (or willing to do so)
 - Stimulate adoption of FAIR4RS guidelines
 - Document & share examples of FAIR4RS adoption (and plans)
- 7. Governance
 - Create communications plan and content that clarifies post-release governance structure

Join one or multiple subgroups via [the form](#)

Next Steps

- Publicize principles
- Via new subgroups:
 - Develop curriculum and training
 - Develop and encourage tooling to support applying principles
 - Encourage adoption, and highlight successful examples
 - Consider future governance of community and change processes
- Community work around gaps that prevent adoption
- Define metrics to measure adoption
 - For specific software
 - For principles and scholarly community as a whole

Get involved!

- Join the RDA [group](#) and be part of the mailing list
- Join one or multiple subgroups via [the form](#)
- Come to [events](#)
- Follow the [steering committee meeting minutes](#)
- Say 'Hi' on the [gitter channel](#)
- Visit and read the publications on [Zenodo](#)
- Review the bibliography collected on [Zotero](#)

All this information is detailed on the [community engagement channels page](#)

Thank you!

Questions?

Keep in touch: morane@softwareheritage.org
@moraneottilia, @SWHeritage
<https://www.fairsfair.eu/fairsfair-newletters/>
<https://www.softwareheritage.org/newsletter/>