

FAIR 4 Research Software (FAIR4RS)

RDA

Morane Gruenpeter, 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

23rd September, 2021

Today's goals

[10.5281/zenodo.5524726](https://doi.org/10.5281/zenodo.5524726)

- Introduction
 - Software in Research
 - Software in RDA
 - Software in a FAIR ecosystem
- Present the FAIR4RS WG activities
- Review FAIR principles for research software
- Challenges
- Future steps

FAIR4RS WG - FAIR Principles for Research Software

Introduction to Research Software

Morane Gruenpeter

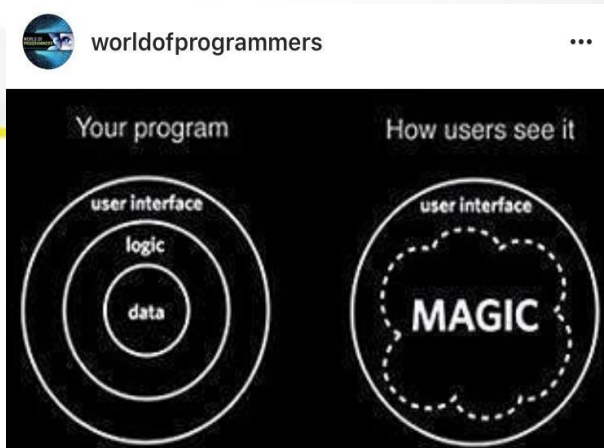
These slides [10.5281/zenodo.5524726](https://doi.org/10.5281/zenodo.5524726)

Question 1

Do you deal with software in your daily work?

- Yes, I use it
- Yes, I develop it
- Yes, I design it
- Yes, I test it
- Yes, all of the above
- No or Not yet

What is software?



Software as a concept

- **project** or entity
- the **community** around the project
- the software **idea** / algorithms / solutions

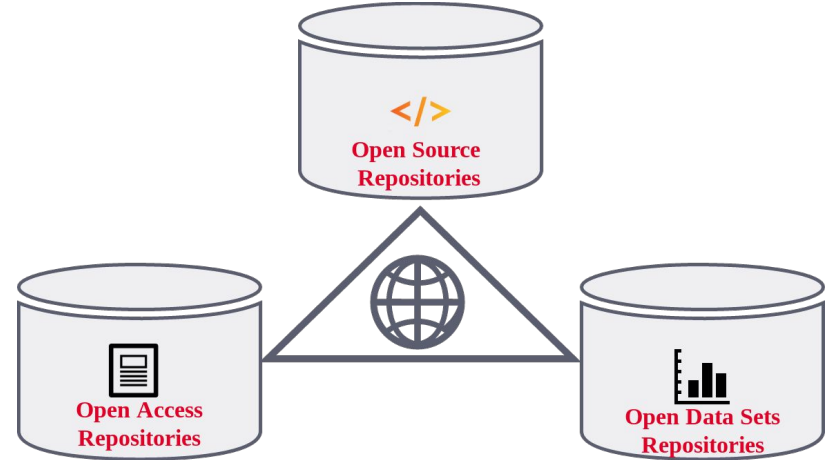
Software artifact

- **source code** form
 - for each version and revision/commit
- **binaries/ executables/ containers** produced (for different environments)

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
Gruenpeter, Software Heritage CC-BY 4.0 2019*

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

Question 2

What kind of stakeholder are you in the scholarly ecosystem?

- a researcher
- a research software engineer
- a publisher
- a citation manager
- a funder
- an evaluator
- a registry admin
- a curator
- or some other role (You can add your role in chat)

Software in RDA

- Software groups:
 - RDA [Software Source Code IG](#) (SSC IG)
 - RDA, ReSA and FORCE11 [FAIR for Research Software Working Group](#) (FAIR4RS WG)
 - RDA & FORCE11 [Software Source Code Identification WG](#) (SCID IG)
 - Output published in September 2020
 - FORCE11 [Software Citation Implementation Working Group](#) (SCIWG)
 - Ongoing WG about software citation

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

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

Chronology - Software in RDA

BOF RDA P9, [Barcelona April 2017](#)

motivations => 60 participants

RDA P10, [Montreal September 2017](#)

SSC IG session- motivations, survey of ontologies, metadata use cases

RDA P11, [Berlin March 2018](#)

SSC IG session- started the idea for a dedicated identification WG

RDA P13, [Philadelphia April 2019](#)

SSC IG session- FAIR for Software Source Code

launch of the **SCID WG**

FORCE2019 (SCIWG), [Edinburgh October 2019](#)

full day hackathon on **Research Software**
(motivated the [CodeMeta generator](#))

RDA VP15, [Australia March 2020](#) (Virtual)

SSC IG session- Open discussion about the creation of a new group, the **FAIR4RS WG**

RDA VP16, [Costa Rica November 2020](#) (Virtual)

First FAIR4RS session

RDA VP17, [Edinburgh April 2021](#) (Virtual)

Getting feedback after the first consultation of the FAIR4RS

Motivation - *Software is not just another type of data*

- FAIR Principles, are intended to apply to **all digital objects** ([Wilkinson et al. 2016](#))
- We focus on the **adaptation and adoption** of the FAIR principles to research software

Recommendation n°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 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](#))

FAIR4RS WG

Carlos Martinez

These slides [10.5281/zenodo.5524726](https://doi.org/10.5281/zenodo.5524726)

Question 3

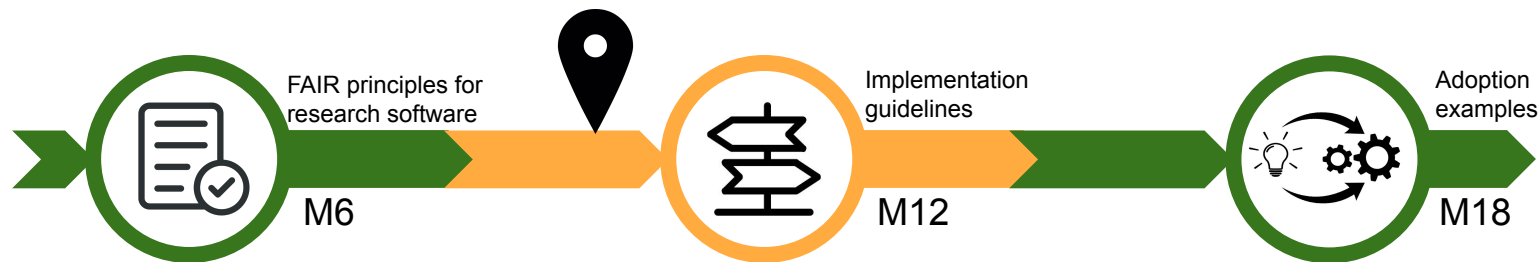
Have you heard about the FAIR4RS WG or read the RDA/ReSA/FORCE11 FAIR4RS principles?

- yes! I even read and commented during the community review
- yes, only heard about it
- not yet :)

Introduction #FAIR4RS

Endorsed by RDA
Sep 2020

- A joint **RDA** Working Group, **FORCE11** Working Group, and Research Software Alliance (**ReSA**) Taskforce.
- Coordinating of a range of existing community-led discussions on:
 - How to define and effectively apply FAIR principles to research software,
 - How to achieve adoption of these principles.



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

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**

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)

Sprints - Drafting the principles

- Small task force
- Five meetings in **April 2021**
- Discussed:
 - feedback from the first consultation
 - Issues from the subgroup work

“maximize the added-value gained by contemporary, formal scholarly digital publishing”

“ensure transparency, reproducibility, and reusability”

The foundational principles of Findable, Accessible, Interoperable, and Reproducible may need to be reinterpreted to ensure that they are applicable to software.

Key challenges:

- Finding the balance between general, more abstract, principles that capture the ethos of FAIR vs specific principles that point to the means of implementation
- Sticking closely to the FAIR data principles or reinterpreting the FAIR ethos for software

Key assumptions:

- Application of the FAIR principles is the responsibility of the owner (often the creator) of the software, not the users
- Principles can be applied to any software used in research
- Software has a wide range of useful lifetimes, and findability, accessibility, interoperability and reusability will degrade over time. This shouldn't be an excuse not to apply FAIR to software

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:

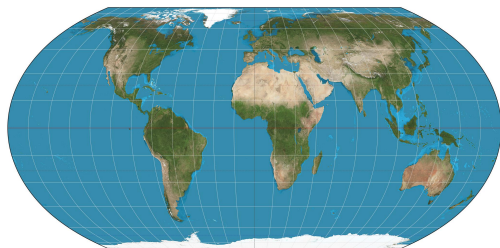
US, UK, NZ, CA, NL, FR,
MX, “Africa”, “Nordics”,
and more!

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>

FAIR Principles for Research Software

Neil P. Chue Hong

These slides [10.5281/zenodo.5524726](https://doi.org/10.5281/zenodo.5524726)

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

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

Findable

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.

Accessible

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

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.

Reusable

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>.

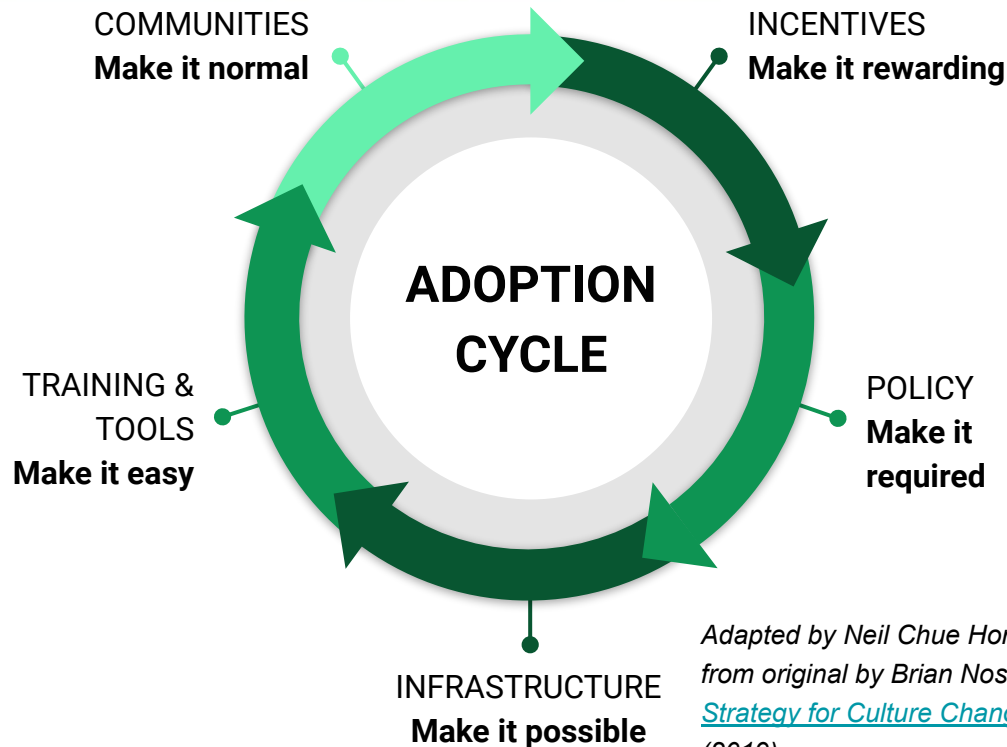
Who is responsible for FAIR software?

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)

Question 4

Should Research Software be FAIR?

- a. Yes, all software used for research should be FAIR
- b. Yes, but only software that is created during the research process
- c. Maybe, depending on the context
- d. No, FAIR is not appropriate for software
- e. Not sure, I'm new to this subject



[Figure: FAIR in a nutshell. Image: ARDC 2018 - CC-BY 4.0.](#)

Challenges

Daniel S. Katz

These slides [10.5281/zenodo.5524726](https://doi.org/10.5281/zenodo.5524726)

Metadata and identifier authority

- Research software needs unique identifiers & associated metadata
 - How are these identifiers created?
 - How is metadata created, stored and maintained?
- Intrinsic metadata (e.g., codemeta.json file in source code repo) is guaranteed to be controlled by the authors but must be exposed to make the software findable
- Extrinsic metadata (e.g., with persistent identifiers) can make software findable but is controlled by external authority

Metadata vocabularies and metadata properties

- Today no community agreement on which vocabularies should be used
- Vocabularies used by package managers to describe software don't capture metadata about research
- There are relatively few discipline-specific vocabularies that capture metadata about software development and usage
- Establishing metadata vocabularies/standards is an intensive process for which resources are limited

Software identifiers

- Today no community agreement on the best identifiers for software
 - Even for specific use cases such as giving software authors credit
- These identifiers are mostly independent and not clearly interoperable
- This could be partly addressed through a **community endorsement process**, in one or more relevant communities

Identification target

- Today no community agreement on what a software identifier should refer to
 - E.g. for open source software, for commercial software, for a container, for a service, etc
- Discussed in FAIR4RS Principles when talking about granularity and versions; also related to the idea of a software concept, which is the set of all specific versions of that software
- Other work, such as ([Hata et al., 2021](#)), has highlighted challenges related to the linkage of scientific knowledge and software artifacts

Software structure complexity

- Software often a complex object
 - Made of other software, documentation, data and metadata, with versions that may change at different rates
- How is this dealt with?
- Where should the FAIR4RS Principles be applied, and where should other interpretations of the FAIR Guiding Principles be applied?
- What should have identifiers, and how should relationships between them be described to be FAIR?
- Here, experience from applying the FAIR Guiding Principles to complex data collections may provide solutions

FAIRness of related research objects

- There is still debate over whether FAIR is recursive, i.e. a digital research object is only “**fully FAIR**” if the objects it builds on are also FAIR
- However, even if just applied to data dependencies, this would restrict the implementation of FAIR4RS Principles as it would require measurable, actionable guiding principles to be applied down the complete dependency stack
- This would ultimately be intractable as the authors of the software would not have responsibility for making the dependencies FAIR

Definitions

Accessibility

- Software engineering has a different, well-understood definition of software accessibility
- Even if the meaning used in the FAIR4RS Principles is well-defined and scoped, it may lead to confusion and mean the principle is not well-understood across all domains.

Reusability

- Software engineering says for software to be reusable it should also be maintainable and dependable (able to be built on for other purposes)
- This may be captured in R3, around domain-relevant community standards, but may also require additional clarification to avoid confusion or the proliferation of many competing sets of “added letters” to FAIR4RS related to other qualities.

Reminder: FAIR is not the end goal

Dan's opinion: goals for research software & research that relies on it

- All research software is open
- All research software is high-quality and robust
- **All research software is findable, accessible, and usable & used by others (for their own research), and is cited when it is used**
- All contributors to research software are recognized for their work, with good careers
- All research software is sustained as long as it is useful
- All research is reproducible

Question 5

Which aspect of making Research Software FAIR you think could be most challenging?

- Metadata - knowing how to describe software (and where to do so)
- Identifiers - identifying the software artifact
- Structure - complexity of software structure and extensive dependencies
- FAIRness of related objects
- Definition of accessibility
- Definition of reusability
- At least 3 of the challenges
- All challenges

Next Steps

These slides [10.5281/zenodo.5524726](https://doi.org/10.5281/zenodo.5524726)

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](#)

Acknowledgements

- Working group chairs: Michelle Barker, Leyla Jael Garcia Castro, Neil Chue Hong, Jennifer Harrow, Daniel S. Katz, Carlos Martinez, Paula A. Martinez, Fotis Psomopoulos and Morane Gruenpeter
- All 228 members and contributors 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 of the FAIR4RS WG

Final question

What is the most important item we can provide to push forward FAIR4RS in your institutions?

(please answer in chat)

Thank you for joining

Questions?

These slides [10.5281/zenodo.5524726](https://doi.org/10.5281/zenodo.5524726)

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