FOUR RECOMMENDATIONS TO MAKE RESEARCH CODE VISIBLE

02 March 2022

PRESENTED BY

DR PAULA ANDREA MARTINEZ
0000-0002-8990-1985
Software Project Coordinator
Australian Research Data Commons

DOI https://doi.org/10.5281/zenodo.6320583

Australian Research Data Commons
NCRIS National Research Infrastructure for Australia
An Australian Government Initiative
ARDC is enabled by NCRIS

ARDC RESEARCH SOFTWARE PROGRAM

Agenda has three layers of distinct action to **See**, **Shape** and **Sustain** research software

TIME FOR A
CULTURAL
SHIFT...
WE SHOULD VALUE REPRODUCIBILITY AS MUCH AS # OF PAPERS PUBLISHED
Scriberia

Image from Scriberia for The Turing Way community

SEE - RESEARCH SOFTWARE

Core concern: Sharing/Availability

Authors: Researchers

Primary driver: Research integrity

Eye icons created by Freepik - Flaticon

“Research is now fundamentally connected to software”
“It permeates every aspect of the conduct of research.”

Perkel, 2021 *10.1038/d41586-021-00075-2*

Neil Chue Hong, Director

SSI, Software Sustainability Institute

[We must recognise] “that re-use and value of data can depend on the availability of relevant metadata, algorithms, code, and software, together with information on workflows and the computational environment used to generate published findings...”

2021. OECD Recommendation of the Council concerning Access to Research Data from Public Funding
https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347

RESEARCH SOFTWARE

Source code files, algorithms, scripts, computational workflows and executables that were created in either of two categories:

A. Within a research project as a by-product to do the research, or

B. Through intentional development of a software product for general use in research by one or more projects.

(adapted from Gruenpeter et al. 2021).

DOI https://doi.org/10.5281/zenodo.5504016

GOALS

- Practical suggestions that contribute to making research software and its source code more discoverable, reusable and transparent.
- Discuss the alignment of these recommendations with the FAIR Principles for research software (FAIR4RS).

NOTES

- We are not covering software development best practices.
- The recommendations are technology independent.

FAIR RESEARCH SOFTWARE #FAIR4RS

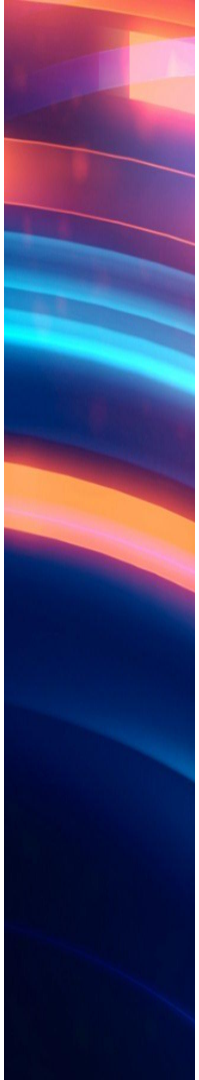
FINDABLE

ACCESSIBLE

INTEROPERABLE

REUSABLE

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



FAIR PRINCIPLES FOR RESEARCH SOFTWARE

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

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

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

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

ABOUT THE RECOMMENDATIONS

Jiménez RC, Kuzak M, Alhamdoosh M *et al.* Four simple recommendations to encourage best practices in research software [version 1; peer review: 3 approved]. *F1000Research* 2017, **6**:876 (https://doi.org/10.12688/f1000research.11407.1)

https://softdev4research.github.io/4OSS-lesson/

DOI https://doi.org/10.12688/f1000research.11407.1

1. Develop source code in a version controlled repository from the beginning of the project

- A software repository or "repo" is a storage location for software
- Version Control System
- Build software “together”
- Private vs Public
- Self-hosted vs web-based, platform independent
- Time is key
- Increases **transparency** through community scrutiny

DOI https://doi.org/10.12688/f1000research.11407.1
https://alternativeto.net/category/developer-tools/version-control-system/

1. **Develop source code in a version controlled repository from the beginning of the project**

alternativeto.net/category/developer-tools/version-control-system

readme.so

1. **Develop source code in a version controlled repository from the beginning of the project**

   REUSABLE

   R: Software is both usable (can be executed)
   and reusable (can be understood, modified, built upon, or incorporated into other software).
   R1.2. Software is associated with detailed provenance.

2. Adopt a licence and comply with the licence of third-party dependencies

- Terms of use, change, contribute, redistribute
- Talk to the Copyright Officer
- choosealicense.com
- Depends on your needs, community standards
- The SPDX license list spdx.org/licenses/
- Builds trust and transparency in how software is created, distributed, and consumed
- ARDC Research Software Rights Management Guide https://doi.org/10.5281/zenodo.5003962

DOI https://doi.org/10.12688/f1000research.11407.1

> “Always license your code. Unlicensed code is closed code.”

Jake VanderPlas, Director of Research in the Physical Sciences, the University of Washington’s eScience Institute. The Whys and Hows of Licensing Scientific Code. *https://www.astrobetter.com/blog/2014/03/10/the-whys-and-hows-of-licensing-scientific-code/*

https://choosealicense.com/no-permission/

2. Adopt a licence and comply with the licence of third-party dependencies

ACCESSIBLE

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

REUSABLE

…can be understood, modified, built upon, or incorporated into other software.

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

R3. Software meets domain-relevant community standards.

DOI https://doi.org/10.12688/f1000research.11407.1

> “Generally when scientists make their code public, they do so because they want it to be free to use and as useful as possible for as many people as possible.”

Jake VanderPlas, Director of Research in the Physical Sciences, the University of Washington’s eScience Institute. The Whys and Hows of Licensing Scientific Code. https://www.astrobetter.com/blog/2014/03/10/the-whys-and-hows-of-licensing-scientific-code/

3. Define clear and transparent contribution, governance and communication processes

- Communication channels and ways to get involved
- End users
- Your collaborators
- Evaluators/ Reviewers
- https://galaxyproject.org/community
- https://softdev4research.github.io/4OSS-lesson/04-contributions/index.html

DOI https://doi.org/10.12688/f1000research.11407.1

“Generally when scientists make their code public…
They want others to not only use it,
but also extend it,
fix bugs,
incorporate it into their own research code,
and thereby make it even more useful to more people.”

Jake VanderPlas, Director of Research in the Physical Sciences, the University of Washington’s eScience Institute. The Whys and Hows of Licensing Scientific Code. https://www.astrobetter.com/blog/2014/03/10/the-whys-and-hows-of-licensing-scientific-code/

3. Define clear and transparent contribution, governance and communication processes

- I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.
- R3. Software meets domain-relevant community standards.

DOI https://doi.org/10.12688/f1000research.11407.1

4. Make software easy to discover by providing software metadata via a community registry

- A Registry stores information about the software
- Facilitates discoverability
- Increases the visibility of the project
- Increases the chances of collaboration, reuse, and improvement
- https://codemeta.github.io/

DOI https://doi.org/10.12688/f1000research.11407.1

4. Make software easy to discover by providing software metadata via a community registry

- Metadata might include information like the source code location, contributors, licence, version, identifier, references and how to cite the software.
- Exposes metadata in a machine readable format
- https://github.com/NLeSC/awesome-research-software-registries
- Domain specific first
- https://www.software.ac.uk/which-journals-should-i-publish-my-software
- https://citation-file-format.github.io/

DOI https://doi.org/10.12688/f1000research.11407.1

4. Make software easy to discover by providing software metadata via a community registry

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

F2. Software is described with rich metadata.

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

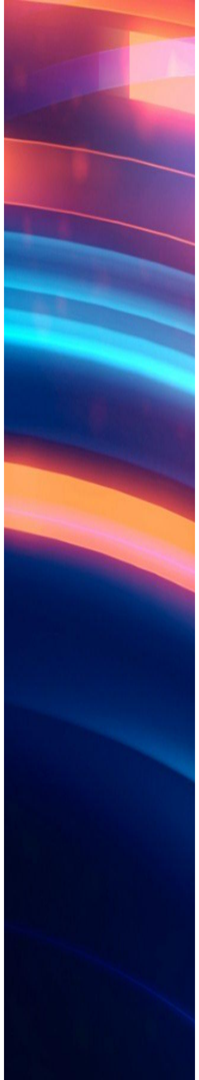
F4. Metadata are FAIR, searchable and indexable.

DOI https://doi.org/10.12688/f1000research.11407.1

4. Make software easy to discover by providing software metadata via a community registry

A1. Software is retrievable by its identifier using a standardized communications protocol.
A2. Metadata are accessible, even when the software is no longer available.
I2. Software includes qualified references to other objects.
R1. Software is described with a plurality of accurate and relevant attributes.
R2. Software includes qualified references to other software.

DOI https://doi.org/10.12688/f1000research.11407.1



FOUR RECOMMENDATIONS TO MAKE RESEARCH CODE VISIBLE

1. Develop source code in a version controlled repository from the beginning of the project
2. Adopt a licence and comply with the licence of third-party dependencies
3. Define clear and transparent contribution, governance and communication processes
4. Make software easy to discover by providing software metadata via a community registry

DOI https://doi.org/10.12688/f1000research.11407.1

“

[Recommends] “4. Take steps to make research data and other research-relevant digital objects from public funding understandable and re-usable in the long term, including through the provision of high quality human-readable, machine-actionable, and open metadata and adequately maintained and supported bespoke algorithms, code, software, and workflows essential for re-use of data as free and open source.”

2021. OECD Recommendation of the Council concerning Access to Research Data from Public Funding https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347

FOR AUTHORS

- Increase discoverability and **visibility** of research software
- **Engages** authors and user **communities**
- Provides **recognition** for contributors
- Builds **trust** among users

DOI https://doi.org/10.12688/f1000research.11407.1

HOW CAN YOU ENABLE THESE RECOMMENDATIONS?

1. Endorsement: agreeing to support the recommendations without a formal process for implementation.
2. Promotion: actively publicising and incentivising the recommendations within and beyond the organisation.
3. Compliance to formally implement the recommendations within the organisation, with ongoing monitoring and public reporting if possible.

DOI https://doi.org/10.12688/f1000research.11407.1

PRESENTED BY

Dr Paula Andrea Martinez

Software Project Coordinator,
Software *Program, ARDC*

✉ paula.martinez@ardc.edu.au

#ResearchSoftwareAU #FAIR4RS

#VisibleResearchSoftware

in https://www.linkedin.com/in/pambio/

https://doi.org/10.5281/zenodo.6320583

Subscribe to the **ARDC CONNECT** newsletter

THANK YOU

ardc.edu.au

contact@ardc.edu.au

+61 3 9902 0585

@ARDC_AU

Australian-Research-Data-Commons

ARDC
Australian Research Data Commons

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

ARDC is enabled by NCRIS