Title: A Survey on Adoption Guidelines for the FAIR4RS Principles

DOI: https://doi.org/10.5281/zenodo.6374598
March 22, 2022

Authors

Paula Andrea Martinez, Australian Research Data Commons, Research Software Alliance, 0000-0002-8990-1985
Michelle Barker, Research Software Alliance, 0000-0002-3623-172X
Alexander Struck, Cluster of Excellence Matters of Activity, 0000-0002-1173-9228
Leyla Jael Castro, ZB MED Information Centre for Life Sciences, 0000-0003-3986-0510
Christopher Erdmann, American Geophysical Union, 0000-0003-2554-180X
Daniel Garijo, Universidad Politécnica de Madrid, 0000-0003-0454-7145
Sandra Gesing, University of Illinois Discovery Partners Institute, 0000-0002-6051-0673
Axel Loewe, Karlsruhe Institute of Technology, 0000-0002-2487-4744
Javier Moldón, Instituto de Astrofísica de Andalucía (IAA-CSIC), 0000-0002-8079-7608

Peer Review

The FAIR4RS WG community were invited to review this document before its publication. Here are those who have reviewed the document.
Daniel S. Katz, University of Illinois at Urbana-Champaign, 0000-0001-5934-7525
Alejandra N. Gonzalez-Beltran, United Kingdom Research and Innovation, 0000-0003-3499-8262
Ilian T. Todorov, United Kingdom Research and Innovation, 0000-0001-7275-1784
Esther Plomp, Delft University of Technology - Faculty of Applied Sciences, 0000-0003-3625-1357
Nadica Miljković, University of Belgrade - School of Electrical Engineering, 0000-0002-3933-6076
Susanna-Assunta Sansone, University of Oxford; RDA/Force11 FAIRsharing WG, 0000-0001-5306-5690

Introduction

Research software (as defined by Gruenpeter et al., 2021) plays a crucial role in science by supporting, processing or delivering results in most scientific areas, including geosciences (Peckham et al., 2013), computational biology (Prlić et al., 2012), and high-energy physics (The HEP Software Foundation et al., 2020). The FAIR[1] (Findable, Accessible, Interoperable and Reusable) Guiding Principles for scientific data management (Wilkinson et al., 2016) stimulated the definition of FAIR principles for research software, which are referred to as the FAIR4RS Principles (Chue Hong et al., 2022). The FAIR4RS Principles are important for the research community as they consider the specific characteristics of software — such as its executability, composite nature, continuous evolution and versioning.

The provision of adoption guidelines will help implement the FAIR4RS Principles. The FAIR4RS Working Group formed subgroups to address milestones (as contained in the case statement), including a subgroup on adoption guidelines, FAIR4RS Subgroup 5. This document provides an overview of the resources identified and reviewed by this subgroup, along with an analysis of gaps, in order to facilitate the adoption of the FAIR4RS Principles. Emerging resources or projects are also documented, and suggestions are made on how to further advance this area.

We expect this survey and gap analysis will be useful to a range of stakeholders seeking resources on how to implement the FAIR4RS Principles, including members of the Research Software Engineering (RSE) and research data management communities, as well as funding bodies, editorial boards of journals and conferences, and research institutions.

Summary of resources

A list of 30+ online resources have been identified and curated by the FAIR4RS Subgroup 5: Adoption Guidelines (Martinez et al., 2022). These resources are available in the Supplementary materials, and the list is open for additions by the community. This list reflects the wide spectrum of global contributions supporting the implementation of the FAIR Principles, particularly regarding research software. This list is a snapshot of currently available resources, although we expect that new resources will become available in the future and that the contents of the current list will evolve. It is important to note that most of these resources precede the definition of the FAIR4RS Principles; however, these still support their implementation. The resources were manually collected, analyzed, and categorized according to their type: guidelines, tools, metadata schemas and registries/repositories. For each resource detail is also provided on which of the FAIR4RS Principles that the resource supports.

[1] FAIR stands for findability, accessibility, interoperability and reusability and does cover explicitly aspects related to openness, reproducibility, transparency, interpretability, good software development practices or software quality.

Most of the resources fall into the ***guidelines*** category and can be further split into two groups: (i) those tailored to FAIR4RS, i.e., focusing on the principles and some steps that can be taken in order to achieve them, and (ii) those covering broader aspects around research software, including development cycle, good practices, and quality. In the latter group, FAIR might not be explicitly mentioned but some FAIR4RS aspects are still covered as part of the best practices for research software, e.g., assigning unique identifiers to releases or including citation information.

The second most common category is ***tools***. Some of these resources extend beyond the implementation of FAIR4RS, for example, focusing exclusively on citation (e.g., formalizing citation formats). Others cover different aspects of the FAIR4RS Principles, either via checklists or (semi)automated means. Many of these tools identify and extract metadata from a code repository while others help create and validate citation files. *Software citation* is a common theme among the guidelines and tools gathered by the group, which reflects the growing interest in acknowledging scientific software contributions.

The third group is ***metadata schemas***, with only three resources, which are designed to formalize software metadata at different levels of granularity or domain requirements. Two of the schemas are aligned to Schema.org, a community-based project promoting schemas for structured metadata, and the third is FAIRsharing where people can find and add metadata standards.

The fourth category includes both ***software registries and repositories***. Although these resources may not directly mention the FAIR4RS Principles, they are fundamental for software findability (indeed, the fourth element of F refers to the software metadata being registered or indexed in a searchable resource) and they also commonly offer elements that are related to e.g., unique identification, rich metadata, citation and search functionality.

Data collection

This subgroup initiated a crowdsourcing effort to identify relevant resources. All members had the opportunity to provide and describe existing FAIR research software guidelines and tools. During the first two months of the subgroup operation in 2021, subgroup participants (referred to as data providers) added resources to an online spreadsheet. Data providers were encouraged to list resources that they were aware of, authored, or were supported by their institutions. Subsequently, the subgroup organized virtual calls to discuss the resources, their descriptions and the categorization. Over the next two months each data provider added descriptions to resources they were familiar with. This meant that some resources gained descriptions from different data providers. Before the completion of the list, the subgroup leads checked the list and cleaned it (removing items that lacked information or providing complementary information). The resulting list, as contained in the Supplementary material, is the first crowdsourced list of its type.

Alignment with the FAIR4RS principles

Most of the resources address **Findability**, with assignment of persistent identifiers and release versioning being the areas of the FAIR4RS principles guidance is most often given. Some guidelines and tools include deposition or archiving in registries, which is related to F4 “Software is registered or indexed in a searchable resource”. The **Accessibility** principle best covered by the resources is A2 “Metadata are accessible, even when software is no longer available”. **Interoperability** aspects are mostly covered by recommendations regarding requirements, dependencies, testing, and commonly used formats. However, coverage of Interoperability is among the lowest, with less than a third of resources including information related to this. Resources regarding **Reusability** commonly refer to adding licenses and citation (part of the provenance) information.

Licensing and citation of software has been encouraged in recent years by initiatives arising from the FORCE11 Software Citation Working Group. Judging by the number of resources that explicitly mention licensing and citation, good progress has been made in this area. In summary, this is how the FAIR4RS Principles are addressed in the collected resources:

- **Findability**: Usually emphasize the use of a persistent identifier, metadata, and release versioning.
- **Accessibility**: Mostly focus on deposition or archiving in registries.
- **Interoperability**: Only a few resources focus on areas such as dependencies, testing and interchangeable metadata.
- **Reusability**: Commonly refer to using licenses and adding citation information.

Gaps

Most of the aforementioned resources date from before the FAIR4RS initiative and do not cover all aspects of the FAIR4RS Principles. This sub group identified the following gaps where resources are lacking, and where potential community efforts can contribute towards guidance, training, and tools.

Finding existing software: with regards to Findability, a number of guidelines/tools describe how to publish software, and how to make it discoverable, but they lack explicit methods for finding software by others. Related work from Hucka & Graham (2018) shows that researchers tend to find software tools by using search engines, recommendations from colleagues, searching the literature, or by browsing surveys compiled by experts. Hence, learning how to search for relevant software is a skill that should be considered in the future (see RDA - Skills and training curriculums to support FAIR research software IG). Building such skills will potentially lead to greater reuse of software and collaboration on research software sustainability.

Another challenge in the Findability of software is that it is not always clear what is shared in public repositories. Software is often mislabeled as data or as a mixed type, particularly when

depositing both software and data. This ultimately leads to misrepresentation of software (for instance, as data) in databases and their search systems. In other words, search systems do not accurately reflect the software results returned. This is another potential gap to be addressed by registries, tools, and guidance, to enable authors to receive clarifying information on whether to share software separately and/or how to describe mixed types deposits.

The previous point links to the identified **infrastructure gaps**: in an ideal scenario, a user could use one publication platform to search for both code and metadata, and filter for required tasks, i.e., research domain, license, and so on. Currently the software repositories landscape is scattered with platforms offering varying functionality (e.g., deposits of research software as source code, binaries, containers, and with different metadata schemas). While some *collaboration* platforms allow source code search, other *publication* platforms limit search functionality to just metadata. An open issue remains on performing meaningful searches for not only code functionality but specific metadata information such as licensing, software version, software type, related objects, documentation availability, domain standards, programming language, etc.

New resources

As the dissemination of the FAIR4RS Principles continues, new resources for their adoption will continue to become available. Some of those upcoming resources or projects related to adopting the FAIR4RS Principles include.

1. Initiatives are underway under the auspices of the FAIR4RS Roadmap led by the Research Software Alliance. The FAIR4RS Roadmap aims to identify key stakeholders in areas arising from the application of the FAIR principles to research software, to guide strategic planning and investment. A range of initiatives have now been formed to progress particular elements of the Roadmap, and the following are likely to create adoption guidelines to suit their audiences:

- Life Sciences Working Group to support the implementation of the FAIR4RS Principles in the life sciences community, with ELIXIR providing a pilot use case.
- FAIR for Virtual Research Environments (VREs) Working Group will enable coordination between existing communities working with VREs, science gateways, platforms and virtual labs, to define what it means for VREs to be and enable FAIR, and provide guidance to VRE developers in achieving this.

2. The FAIR4RS Working Group, via a different subgroup, also identified early adopters and organizations to promote the quality of research software. The actions being taken are diverse and include examples of training, infrastructure, policies and metrics. Details are provided in the report on “Adoption Support” (Martinez-Ortiz et al., 2022).

3. The European Open Science Cloud (EOSC) builds on existing infrastructure and services supported by the European Commission and aims to facilitate a multi-disciplinary environment where researchers can publish, find and re-use data, tools and services, enabling them to better conduct their work. The EOSC Infrastructures for Quality Research Software Task Force, formed in 2021, will identify standard-based best practices for development of quality research software. Standardization efforts on metadata and identifiers for research software source code are within the scope of the group. Their main objectives are to 1) foster rapid implementation of the actionable recommendations of the EOSC scholarly infrastructures for research software (SIRS) report (European Commission, 2020) and 2) foster the adoption of best practices that may lead to producing better research software, contributing to economic and organizational sustainability of research software development, and improving reproducibility of research results. A complementary effort from the EOSC FAIR Metrics and Data Quality Task Force will work to define indicators to measure the FAIRness of digital objects, also to ensure that FAIR evaluation/assessment/assisting tools deliver consistent results.

4. Work is also underway to define FAIR4RS metrics to facilitate assessment of the 'FAIRness' of research software. The "Fair Metrics for FAIR Software" project from the Open Science Fund by the Dutch Research Council (NWO) will evaluate suggested FAIRness metrics for software while focusing on their incentivization potential. To foster the cultural change towards open science, it is vital that such metrics incentivize software-developing researchers of all disciplines to create 'FAIRer' software. The project will develop recommendations for the use of FAIRness metrics and automated assessment tooling to enable easier and transparent FAIRness assessment (Open Science Fund Applications submitted in 2021).

5. The Australian Research Data Commons is creating a tool for research software outputs, similar to the FAIR Data Self assessment Tool. This tool has been designed to be used by a broad range of stakeholders to explore what steps can be taken to make research software (more) FAIR. This resource lists the FAIR4RS Principles with additional explanatory information, and a list of possible ways to address the FAIR4RS Principles.

Suggested next steps

After the completion of the FAIR4RS Working Group in mid-2022, further work will be needed to widen the scope of resources available to ensure a wider adoption of the FAIR4RS Principles. We believe the following steps should be taken by the community in order to address the identified gaps:

Resource coverage extension: While the authors have undertaken significant effort to compile a comprehensive list of resources, the scientific community is likely to develop new initiatives to promote and ease the adoption of the FAIR4RS Principles. Hence, we have opened the collected

resource list for additional contributions by the community (Martinez et al., 2022). Our expectation is that new resource creators will register their resource with a summary as part of the list in order to increase awareness in the community.

Finding existing software: Learning how to find software is a skill that should be learned by researchers in order to minimize duplication of efforts. Efforts like the RDA - Skills and training curriculums to support FAIR research software IG are needed to address this issue, while also increasing software reuse in the process. There is also potential to improve Findability by technological means. For example, research-focused search engines (like base-search.net using Document Type: software) could index source code instead of just metadata. Registries (like re3data.org) could include more platforms like swMATH.org and describe code repositories with software-related metadata. Domain specific and institutional repositories could include a specific category for software. We envision that repositories holding research software will reinforce the FAIR4RS Principles adoption, and use software specific metadata schemas, licenses and standards to cover more aspects of research software.

Infrastructure: The heterogeneity of platforms for finding, describing and versioning software are tailored towards different users, use cases and domains. Additional work is needed towards the adoption of common metadata standards, in order to ease cross-platform interoperability, software search and evaluation. The work led by the Consortium of Scientific Registries and Repositories (SciCodes) (Montail et al. 2020) is a first step in this direction, promoting FAIR and metadata interoperability among registry developers, editors and maintainers.

Acknowledgements

Michelle Barker's contributions to this work was supported by the Wellcome Trust grant 222436/Z/21/Z.

Alexander Struck acknowledges the support of the Cluster of Excellence Matters of Activity. Image Space Material funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2025 – 390648296.

Daniel Garijo's work is supported by the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with Universidad Politécnica de Madrid in the line Support for R&D projects for Beatriz Galindo researchers, in the context of the V PRICIT (Regional Programme of Research and Technological Innovation)

Javier Moldón acknowledges financial support from the State Agency for Research of the Spanish MCIU through the "Center of Excellence Severo Ochoa" award to the Instituto de Astrofísica de Andalucía (SEV-2017-0709) and from the grant RTI2018-096228-B-C31 (MICIU/FEDER, EU)

References

1. 2022. *Open Science Fund Applications submitted in 2021*. 1st ed. [ebook] NWO, p.67. Available at: <https://www.nwo.nl/sites/nwo/files/media-files/Open%20Science%20Fund%20applications%20submitted%20in%202021.pdf> [Accessed 11 February 2022].
2. 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: https://www.doi.org/10.15497/RDA00068
3. European Commission, Directorate-General for Research and Innovation, *Six Recommendations for implementation of FAIR practice by the FAIR in practice task force of the European open science cloud FAIR working group*, Publications Office, 2020, https://data.europa.eu/doi/10.2777/986252
4. European Commission, Directorate-General for Research and Innovation, Scholarly infrastructures for research software : report from the EOSC Executive Board Working Group (WG) Architecture Task Force (TF) SIRS, Publications Office, 2020, https://data.europa.eu/doi/10.2777/28598
5. Gruenpeter, M., Katz, D. S., Lamprecht, A.-L., Honeyman, T., Garijo, D., Struck, A., … Sinha, M. (2021). *Defining Research Software: a controversial discussion* (Version 1). Zenodo. Retrieved from https://doi.org/10.5281/zenodo.5504015
6. Hucka, M., & Graham, M. J. (2018, July). Software search is not a science, even among scientists: A survey of how scientists and engineers find software. *Journal of Systems and Software*. Elsevier BV. http://doi.org/10.1016/j.jss.2018.03.047
7. Martinez, P. A., Struck, A., Castro, L. J., Garijo, D., Loewe, A., Gesing, S., … Sansone, S.-A. (2022). A Survey on Adoption Guidelines for the FAIR4RS Principles: Dataset (Version 1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.6375540
8. Martinez-Ortiz, C., Katz, D.S., Lamprecht, A.L., Barker, M., Loewe, A., Fouilloux, A., Wyngaard, J., Garijo, D., Moldon, J., Castro, L.J., Wheeler, D., Albers, J., & Lee, A. (2022). FAIR4RS: Adoption support. Zenodo. https://doi.org/10.5281/zenodo.6258366
9. Monteil, A., Gonzalez-Beltran, A., Ioannidis, A., Allen, A., Lee, A., Bandrowski, A., ... & Morrell, T. (2020). Nine best practices for research software registries and repositories: A concise guide. arXiv preprint arXiv:2012.13117.
10. Papers with Code Portal. 2021. [online] Available at: <https://portal.paperswithcode.com> [Accessed 16 February 2022].
11. Peckham, S. D., Hutton, E. W., & Norris, B. (2013). A component-based approach to integrated modeling in the geosciences: The design of CSDMS. Computers & Geosciences, 53, 3-12. https://doi.org/10.1016/j.cageo.2012.04.002
12. Prlić, A., & Lapp, H. (2012, November 29). The PLOS Computational Biology Software Section. *PLoS Computational Biology*. Public Library of Science (PLoS). http://doi.org/10.1371/journal.pcbi.1002799

13. Research Data Alliance. 2022. *FAIR for Research Software (FAIR4RS) WG*. [online] Available at: <https://www.rd-alliance.org/groups/fair-research-software-fair4rs-wg> [Accessed 11 February 2022].
14. Research Data Alliance. 2020. *FAIR4RS Case Statement*. [online] Available at: <https://www.rd-alliance.org/node/69317/case-statement> [Accessed 11 February 2022].
15. Schema.org. 2022. *Schema.org - Schema.org*. [online] Available at: <https://schema.org> [Accessed 11 February 2022].
16. Research Data Alliance. 2022. *Skills and training curriculums to support FAIR research software IG*. [online] Available at: <https://www.rd-alliance.org/groups/skills-and-training-curriculums-support-fair-research-software-ig> [Accessed 11 February 2022].
17. The HEP Software Foundation., Albrecht, J., Alves, A.A. *et al.* A Roadmap for HEP Software and Computing R&D for the 2020s. *Comput Softw Big Sci* **3,** 7 (2019). https://doi.org/10.1007/s41781-018-0018-8
18. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016, March 15). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. Springer Science and Business Media LLC. http://doi.org/10.1038/sdata.2016.18

Supplementary materials

This survey is supplemented by a dataset which can be downloaded and cited:

Martinez, Paula Andrea, Struck, Alexander, Castro, Leyla Jael, Garijo, Daniel, Loewe, Axel, Gesing, Sandra, Barker, Michelle, Chue Hong, Neil, Erdmann, Christopher, Martinez-Ortiz, Carlos, & Sansone, Susanna-Assunta. (2022). A Survey on Adoption Guidelines for the FAIR4RS Principles: Dataset (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6375540

The dataset is a list of 30+ online resources that have been identified and curated by the FAIR4RS Subgroup 5: Adoption Guidelines. The list is open for additions by the community via comments directly to this link.