



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Introducing the FAIR Principles for research software

**Citation for published version:**

Barker, M, Chue Hong, NP, Katz, DS, Lamprecht, A-L, Martinez Ortiz, C, Psomopoulos, F, Harrow, J, Garcia Castro, LJ, Gruenpeter, M, Martinez, PA & Honeyman, T 2022, 'Introducing the FAIR Principles for research software', *Scientific Data*, vol. 9, no. 1, 622. <https://doi.org/10.1038/s41597-022-01710-x>

**Digital Object Identifier (DOI):**

<https://doi.org/10.1038/s41597-022-01710-x>

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Early version, also known as pre-print

**Published In:**

Scientific Data

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## 1. Introduction

In 2016 the publication of “The FAIR Guiding Principles for scientific data management and stewardship”<sup>1</sup> supported a vision where valuable scientific outputs are made ‘FAIR’ by becoming more Findable, Accessible, Interoperable and Reusable. From the outset, the FAIR Guiding Principles were intended to be applicable to many kinds of digital assets. Increased understanding of the importance of research software in research has catalysed application of the FAIR Guiding Principles to this type of digital asset.

Community-endorsed FAIR principles for research software have now been developed by the FAIR for Research Software (FAIR4RS) Working Group (WG), which was jointly convened by the Research Software Alliance (ReSA), Future Of Research Communications and E-Scholarship (FORCE11), and the Research Data Alliance (RDA). This milestone reflects the maturation of the research community in understanding the benefits of having FAIR research software, and coming together as the FAIR4RS WG to achieve this. The FAIR4RS WG is a global and interdisciplinary community whose members have an interest in the application of FAIR principles to research software, such as researchers, software users, developers and maintainers, policy makers, infrastructure support staff, and funders.

The FAIR4RS Principles are relevant to any stakeholder in the research community seeking to increase transparency, reproducibility, and reusability of research. This paper highlights the importance of the FAIR4RS Principles and the positive signals of adoption that demonstrate high levels of community support. The paper is organised as follows. Section 2 details the FAIR4RS Principles, and provides examples of how they can be applied to different types of research software. Section 3 discusses their importance and impact, and provides organisational examples of implementation in practice that readers can learn from and utilise. Finally, Section 4 discusses how the process used to develop the FAIR4RS Principles both leveraged and amplified community understanding of the critical role of research software in maximising research value, leading to the very positive signals of early adoption.

## 2. Results

In this section, each of the FAIR4RS Principles is proposed and explained, as contained in version 1.0.<sup>2</sup> First, each foundational principle (F, A, I and R) is described, followed by the numbered guiding principles used to detail it. It should be noted that the FAIR4RS Principles are aspirational, as with the FAIR Guiding Principles. The community has come together to provide clarity around common goals, by defining simple and research software appropriate goalposts to inform those who publish and/or preserve research software. Contextual information follows, then examples of how the principles can be applied to three types of research software are presented.

Research software is defined by the FAIR4RS WG as including “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.”<sup>3</sup>

## 2.1 FAIR4RS Principles

The FAIR4RS Principles<sup>2</sup> are:

<p><b>F: Software, and its associated metadata, is easy for both humans and machines to find.</b></p>
<p>F1. Software is assigned a globally unique and persistent identifier.</p> <ul style="list-style-type: none"> <li>• F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.</li> <li>• F1.2. Different versions of the software are assigned distinct identifiers.</li> </ul> <p>F2. Software is described with rich metadata.</p> <p>F3. Metadata clearly and explicitly include the identifier of the software they describe.</p> <p>F4. Metadata are FAIR, searchable and indexable.</p>
<p><b>A: Software, and its metadata, is retrievable via standardised protocols.</b></p>
<p>A1. Software is retrievable by its identifier using a standardised communications protocol.</p> <ul style="list-style-type: none"> <li>• A1.1. The protocol is open, free, and universally implementable.</li> <li>• A1.2. The protocol allows for an authentication and authorization procedure, where necessary.</li> </ul> <p>A2. Metadata are accessible, even when the software is no longer available.</p>
<p><b>I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.</b></p>
<p>I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.</p> <p>I2. Software includes qualified references to other objects.</p>
<p><b>R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).</b></p>
<p>R1. Software is described with a plurality of accurate and relevant attributes.</p> <ul style="list-style-type: none"> <li>• R1.1. Software is given a clear and accessible license.</li> <li>• R1.2. Software is associated with detailed provenance.</li> </ul> <p>R2. Software includes qualified references to other software.</p> <p>R3. Software meets domain-relevant community standards.</p>

## 2.2 FAIR research software examples

Three examples of how research software projects implement the principles are provided to increase understanding of how they apply in practice. It should be noted that the application of the FAIR4RS Principles is the responsibility of the owners (who are often the creators) of the software, not the users. However, scholarly infrastructures are needed to provide certain functionalities to apply FAIR to software. The principles can be applied to a wide range of research software, and the examples here are of a command-line tool for a specific task, a collection of scripts and notebooks that form a complex research software product, and

graphical user interfaces to other packages or libraries. The following three examples show how parts of FAIR4RS Principles can be implemented; there are not yet any examples of research software that address all the FAIR4RS Principles as presented above, as they have only been defined recently. For all these examples demonstrating the principles, there is also the possibility that additional actions can be taken to make the software even more FAIR.

- Comet is a command-line tool and desktop application for tandem mass spectrometry sequence database search<sup>4</sup>. It is registered in the bio.tools catalogue of bioinformatics tools, where it has a globally unique and persistent identifier (FAIR4RS Principle: F1), and rich metadata (F2) that includes the identifier (F3) and is searchable and indexable (F4). Comet can be downloaded via the browser following the links provided in the metadata using https (A1). The metadata in bio.tools is independent from the Comet repository, and will stay accessible should the software itself become inaccessible (A2). Comet uses standard data types from the proteomics domain for its input and output data (I1) that are documented in the metadata as functional annotations (I2). The software is licensed under the Apache 2.0 open source licence, and the publicly accessible project repository on GitHub includes detailed information about its development (R1). The code includes dependencies to external software packages, such as Thermo Scientific's MSFileReader library (R2).
- PuReGoMe is a project aimed at understanding Dutch public sentiment during the COVID-19 outbreak period by analysing real-time Twitter data<sup>5</sup>. It provides a collection of Python scripts and Jupyter notebooks for this purpose. PuReGoMe has a (versioned) DOI from Zenodo (F1) and is registered in the Research Software Directory that captures the most relevant metadata (F2), including the identifier (F3), in searchable and indexable form (F4). The software can be downloaded from the project repository (A1), while metadata is accessible independently from the registry (A2). PuReGoMe uses standard file formats (e.g., CSV files) for data exchange (I1) and refers to other objects such as websites (I2). The project uses the Apache 2.0 open source licence, and the GitHub repository has detailed records of the development history (R1). The code includes dependencies to other software, such as various Python libraries (R2).

- A third example of a research software artefact deposited in the HAL French national archive is gammaShiny (<https://hal.inria.fr/hal-02979514v1>) gammaShiny is an application that provides enhanced graphical user interfaces for the R gamma package. It is used to process in-situ gamma-ray spectrometry measurements for luminescence dating. gammaShiny has been deposited in the HAL French national archive<sup>1</sup> and it has a persistent globally unique identifier <sup>6</sup> (F1) with the HAL identifier of the metadata record and a SWHID, identifying specifically the software artefact on the Software Heritage<sup>2</sup> universal software source code archive. Thanks to the HAL platform, where a licence is mandatory, we can observe that gammaShiny is under a GNU General Public Licence v3.0 or later (R1). When browsing the code on SWH, you can find a codemeta.json file, identifiable with a SWHID<sup>3</sup>, where other metadata is available including dependencies named in CodeMeta - `softwareRequirements` (R2).

### 3. Discussion

This section discusses the importance and impact of the FAIR4RS Principles in reflecting research community maturation, and provides adoption examples for readers to learn from to facilitate easier implementation.

#### 3.1 Research community maturation

The development of the FAIR4RS Principles is a milestone for the research community in recognizing the increasing value of research software as fundamental and vital to research worldwide. While improving the practice of scholarship is a common goal of the open science, open source, and FAIR communities, and although FAIR research software can improve data usage, and consequently, improve research, research software is only now emerging as a strong focus of FAIRness.

While many of the FAIR Guiding Principles could be directly applied to research software by treating software and data as similar digital research objects, unique characteristics of software (e.g., its executability, composite nature, continuous evolution and versioning), as well as of the ecosystem in which it is developed and shared (e.g., social coding platforms, package management systems), made it necessary to revise and extend the principles to create the FAIR4RS Principles.

An additional benefit of the development of community-endorsed principles is catalysing shared practices. Enabling community participation increases awareness of the challenges faced by different stakeholders and how the principles might address them. The FAIR4RS WG identified

---

<sup>1</sup> <https://hal.archives-ouvertes.fr/>

<sup>2</sup> <https://archive.softwareheritage.org/>

<sup>3</sup> <https://archive.softwareheritage.org/swh:1:cnt:eff9a3c75042d7b994734f1471331d79556cba8d;origin=https://hal.archives-ouvertes.fr/hal-02979514;visit=swh:1:snp:5a2a9f64c4a462daabc1dc891fbd65e12636a242;anchor=swh:1:rev:713f2344685a8b17af36f82e970df0d1d2d494cd;path=/gammaShiny-0.1.0/codemeta.json>

a range of opportunities for future work<sup>7,8</sup> that share an emphasis on the need for increased standardisation of community practices. These areas include metadata and identifier authority, metadata vocabularies and metadata properties, software identifiers, domain-relevant community standards for software and identification targets. This discussion will make it simpler for people wanting to follow the principles to know how to do so.

## **3.2. Adoption**

The FAIR4RS WG has been successful in facilitating initial adoption of its outcomes, reflecting high levels of initial impact. Adoption and implementation of the FAIR4RS Principles will create significant outcomes for many stakeholders, ranging from increased clarity for funders around their own requirements for software investments to guidelines for publishers and research institutions on sharing requirements. The FAIR4RS Principles are also relevant to the larger ecosystem and the stakeholders that support research software (e.g., repositories and registries). A number of groups and organisations are planning to adopt the FAIR4RS Principles, including ELIXIR, Australian Research Data Commons (ARDC), Netherlands eScience Center (NLeSC), and ZB MED, and are at various stages at this point, as described in this subsection. This demonstrates the initial organisation support that the FAIR4RS Principles are receiving, and provides information on policies, guidelines and activities that other organisations could consider implementing.

The RDA Software Source Code Interest Group is the maintenance home for the FAIR4RS Principles. Concerns or queries about the principles can be raised at RDA plenary events organised by this Interest Group, where there may be opportunities for adopters to report back on progress. In two years time, the community will gather again to see if anything needs to be changed.

### **3.2.1. Australian Research Data Commons**

The ARDC is a national facility enabled by the Australian Government's National Collaborative Research Infrastructure Strategy. The mission of the ARDC is to accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality datasets. To support this, the software program at the ARDC is working towards recognition of research data as a first-class output of research. The ARDC National Agenda for Research Software includes activities seeking to make research software outputs more FAIR in Australia<sup>9</sup>. The ARDC is undertaking a number of actions to implement the FAIR4RS Principles, beginning with policy change. ARDC's investment policy lays out expectations for partners to make the outputs of projects receiving ARDC investment more FAIR. ARDC's policy for FAIR outputs is being updated to reference the FAIR4RS Principles as appropriate for software outputs arising from future investment.

In terms of implementation, the ARDC provides national support for adopting the FAIR Guiding Principles through collaborations, activities and guidance materials. This support is expanding to

include the FAIR4RS Principles as well as development of an equivalent guidance for Virtual Research Environments (VREs) through the RDA FAIR4VREs WG. The initial focus for engagement is through the ARDC platforms community, which represents 26 projects developing national platforms infrastructure with a total investment of AUD 58M. This will be followed with activities targeting broader research software communities across Australia. Activity around materials development has already begun. A FAIR research software counterpart is being developed to complement the highly successful ARDC FAIR data self-assessment tool. This new tool will be used to raise awareness of the FAIR4RS Principles, and to help a wide variety of stakeholders identify and order actions to target in adopting the FAIR4RS Principles. Finally, while not a major producer of research software itself, the ARDC will make some of its own software outputs FAIR to demonstrate the effort and impact of doing so.

### **3.2.2. ELIXIR**

ELIXIR coordinates and develops life science resources across Europe so that researchers can more easily find, analyse and share data, exchange expertise, and implement best practices. As a European intergovernmental organisation that is made up of life scientists, computer scientists, and support staff, ELIXIR unites Europe's leading life science organisations to help researchers take advantage of the huge amounts of data produced in life science, to gain new insights into how living organisms work in health and disease. This is achieved by coordinating, integrating and sustaining bioinformatics resources across ELIXIR member states and enabling users in academia and industry to access services that are vital for their research.

Advancing the understanding of life and disease over all the domains in life sciences requires that research data, analysis tools, standards and computational services are FAIR. ELIXIR policy recommends that all research outputs of ELIXIR infrastructure be FAIR, including a strong recommendation to all ELIXIR partners that software developed and supported by ELIXIR be FAIR. This includes standalone tools (such as InterMine), platforms (such as Galaxy) as well as services (such as bio.tools and OpenEBench). Moreover, and in the context of the FAIR assessment and adoption process, of particular interest are services that enable and support FAIRness in research outcomes, such as FAIRsharing. This is an RDA-endorsed resource, recommended by ELIXIR and other organisations across all disciplines. In the context of research software, FAIRsharing covers: (i) the standards used to describe the software metadata; (ii) the standards used as software input and output formats; and (iii) the software code repositories.

Additionally, one of the most recent achievements of the ELIXIR Tools Platform is the ELIXIR Software Management Plan<sup>10</sup>, a low-barrier standard specifically tailored for life science researchers for capturing the life cycle for the software that is produced within the lifetime of a particular project or activity, specifically tailored for life science researchers<sup>10</sup>. As a clear outcome of ELIXIR adopting the FAIR4RS Principles, a roadmap has been put in place to ensure that the ELIXIR Software Management Plan, currently implemented as a standalone tool based on the Data Stewardship Wizard<sup>11</sup>, is fully aligned to them. Specifically, the sections and individual questions of the Software Management Plan will be clearly annotated to the

corresponding FAIR Principles, in order to both increase the awareness of the principles themselves, as well as promote FAIRness for software within ELIXIR.

There are also several training efforts under the ELIXIR Training Platform related to FAIR; FAIR Data Management and Data Stewardship training, FAIR Training, the Terms4FAIRskills initiative, etc. A new set of training materials will be developed within the next two years to increase the use of Software Management Plans and understanding of the underlying FAIR4RS principles.

### **3.2.3. Netherlands eScience Center**

The Netherlands eScience Center is the Dutch national centre of research software expertise, and has an internal strategy for Open Science and Software Sustainability. This document highlights the centrality of software quality, reusability, and adoptability to the eScience Center's vision. The FAIR4RS Principles are a key element of this vision particularly related to software reusability.

The eScience Center is adopting the FAIR4RS Principles by:

- Using the FAIR4RS Principles to support the creation of reusable software as part of its calls and projects. All projects funded by the eScience Center are required to have software management plans.
- Contributing to developing the skills necessary to implement the FAIR4RS Principles through skills and knowledge development activities: digital skills training, development of practical guidelines, etc. Their Digital Skills Programme helps develop some of the necessary skills required to develop software that is as reusable as possible.
- Updating the Five Recommendations for FAIR software tool<sup>12</sup> to better align with the FAIR4RS principles.
- Promoting the principles on relevant policy agendas at a (inter)national level. For example, the eScience Center is collaborating with the Dutch Research Council (NWO) in the creation of national templates for Software Management Plans. The aim of these templates is to provide guidance for Dutch research organisations and scientific communities, as well as individual researchers, on how to organise research software and ensure its sustainability. The FAIR4RS Principles are being considered as a starting point in the development of these templates. The guidance provided by the templates will be, wherever possible, closely aligned with the FAIR4RS Principles.

### **3.2.4. ZB MED**

ZB MED Information Centre for Life Sciences is a German national infrastructure and research centre for information and data in the life sciences. The mission of ZB MED is to ensure the national provision of information and literature in the fields of medicine, health care, nutritional, environmental and agricultural sciences – including the relevant basic sciences and related subject areas – for the purposes of research, teaching and practical application. To support this,



ZB MED follows and encourages an open science policy so that all research outcomes become open to the public, and adhere to the FAIR Guiding Principles and good research management practices. However, data is only one piece of the puzzle. For research to become reproducible and sustainable, connected FAIRification is needed across all the pieces, including data, software, workflows and other research objects.

ZB MED promotes the FAIR Guiding Principles by providing support, advice and training to researchers in this regard. ZB MED is now adopting and promoting the FAIR4RS Principles to better support research software. Researchers at ZB MED already follow some practices related to FAIR and openness, including the recommendations for open source code<sup>13</sup>, combined with the Five Recommendation for FAIR Software tool<sup>12</sup>. Currently, ZB MED is adapting and extending these practices to better align them to the FAIR4RS Principles and the adoption guidelines that the same group is working on. Furthermore, ZB MED plans to extend dissemination activities around FAIR to include the FAIR4S Principles so that more researchers, at national and international level, understand and adopt them.

#### **4. Methods**

The adoption of the FAIR4RS Principles by a range of organizations reflects the extensive consultation undertaken by the FAIR4RS WG with the research community. The FAIR4RS WG has engaged about 500 people (from more than 110 organizations over more than 34 countries) in the development of the principles, including the more than 240 FAIR4RS WG members. “The FAIR4RS team: Working together to make research software FAIR”<sup>14</sup> and the FAIR4RS Community Profile<sup>15</sup> provide details on the approach to community collaboration, showcasing a model for teamwork across the research software community.

The FAIR4RS WG efforts began with subgroups from July 2020 to March 2021 that provided outputs to support the development of the FAIR4RS Principles. This work was brought together in a single report and presented for feedback by the wider FAIR4RS community in March 2021<sup>8</sup>. Draft FAIR4RS Principles were released in June 2021<sup>16</sup> and underwent a formal community review process for one month. The current version of the FAIR4RS Principles, as summarized here, is a cumulative result of this work.

Part of the reason for the FAIR4RS WG’s very high levels of success in community engagement is because it brought together a range of efforts to apply aspects of FAIR to research software since 2017<sup>8</sup>, and because it sought to align with a range of FAIR data efforts. In this way the FAIR4RS WG was able to leverage and amplify existing community momentum to demonstrate the benefits of implementing the principles.

#### **Data Availability**

n/a

#### **Code availability**

n/a

## References

1. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
2. Chue Hong, N. P. *et al.* FAIR Principles for Research Software (FAIR4RS Principles). *Research Data Alliance* (2022) doi:10.15497/RDA00068.
3. Gruenpeter, M. *et al.* *Defining Research Software: a controversial discussion*. <https://zenodo.org/record/5504016> (2021) doi:10.5281/ZENODO.5504016.
4. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: An open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24 (2013).
5. Tjong Kim Sang, E., Schraagen, M. & Wang, S. *PuReGoMe*. (Zenodo, 2020). doi:10.5281/ZENODO.4337242.
6. Frerebeau, N. & Lebrun, B. gammaShiny: ‘Shiny’ Applications for the R Package ‘gamma’. (IRAMAT-CRP2A (UMR 5060), 2020). doi:10.5281/zenodo.4139005.
7. Katz, D. S., Gruenpeter, M. & Honeyman, T. Taking a fresh look at FAIR for research software. *Patterns* **2**, (2021).
8. Katz, D. S., Chue Hong, N. P., Barker, M. & Gruenpeter, M. FAIR4RS WG subgroup community consultation March 2021. (2021) doi:10.5281/ZENODO.4635410.
9. Honeyman, T. & Treloar, A. A National Agenda for Research Software. (2021) doi:10.5281/ZENODO.4940273.
10. Alves, R. *et al.* *ELIXIR Software Management Plan for Life Sciences*. <https://osf.io/k8znb> (2021) doi:10.37044/osf.io/k8znb.
11. Pergl, R., Hooft, R., Suchánek, M., Knaisl, V. & Slifka, J. “Data Stewardship Wizard”: A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning. *Data Science Journal* **18**, 59 (2019).
12. Netherlands eScience Center & DANS. Five Recommendations for FAIR Software. *FAIR Research Software* <https://fair-software.nl/home> (2020).
13. Jiménez, R. C. *et al.* Four simple recommendations to encourage best practices in research software. *F1000Res* **6**, 876 (2017).
14. Katz, D. S., Barker, M., Chue Hong, N. P., Castro, L. J. & Martinez, P. A. The FAIR4RS team: Working together to make research software FAIR. (2021) doi:10.5281/ZENODO.5037157.
15. Woodley, L., Pratt, K., Kobilka, S. & Martinez, P. A. CSCCE Community Profile: The FAIR for Research Software (FAIR4RS) Working Group. (2021) doi:10.5281/ZENODO.5019939.
16. Chue Hong, N. P. *et al.* FAIR Principles for Research Software (FAIR4RS Principles). *Research Data Alliance* (2021) doi:10.15497/RDA00065.