The FAIR4RS team: Working together to make research software FAIR

Daniel S. Katz*, Michelle Barker†, Neil P. Chue Hong‡, Leyla Jael Castro§, Paula Andrea Martinez¶

*NCSA & CS & ECE & iSchool, University of Illinois, Urbana, IL, USA, ORCID: 0000-0001-5934-7525
†Research Software Alliance, Cairns, Australia, ORCID: 0000-0002-3623-172X
‡Software Sustainability Institute & EPCC, University of Edinburgh, Edinburgh, UK, ORCID: 0000-0002-8876-7606
§ZB MED Information Centre for Life Sciences, Cologne, Germany, ORCID: 0000-0003-3986-0510
¶Research Software Alliance, Brisbane, Australia, ORCID: 0000-0002-8990-1985

Abstract—The FAIR For Research Software Working Group (FAIR4RS WG) is leading the research software community in the crucial step of agreeing how to apply the FAIR principles to research software, including defining the principles and adoption guidelines. This group was convened under the Research Software Alliance (ReSA), the Research Data Alliance (RDA), and FORCE11 (the Future of Research Communication and Escholarship). This paper explains how the community collaboration to enable this has been built, as an example of a model for teamwork across the research software community, and which could potentially be used in other fields and domains as well, as this is an interdisciplinary and global effort.

Index Terms—FAIR software, research software, community, collaboration

I. INTRODUCTION

Research software is a significant and vital component of research. It is integral to all stages of current research practices, including data collection and generation, curation, pre- and post-processing, analysis, and modeling. Recognizing software's roles in reproducibility and the need for recognition for those who develop software has energized conversations on what FAIR (Findable, Accessible, Interoperable, Reusable) [1] means for research software [2, 3].

To improve the FAIRness of research software, we are working to develop and apply the FAIR Guiding Principles to such software. Many of the high-level FAIR data principles can be directly applied to research software by treating software and data as similar digital research objects. However, specific characteristics of software, such as its executability, composite nature, and continuous evolution and versioning [4], make it necessary to revise and extend the original FAIR data principles.

The significance of this work is illustrated by the convening of the FAIR For Research Software Working Group (FAIR4RS WG, www.rd-alliance.org/groups/fair-4-research-software-fair4rs-wg) across the Research Data Alliance (RDA), FORCE11 (the Future of Research Communication and eScholarship), and the Research Software Alliance (ReSA). RDA builds the social and technical bridges to enable the open sharing and re-use of data. FORCE11 is a community of scholars, librarians, archivists, publishers and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing.

ReSA's vision is to see research software recognised and valued as a fundamental and vital component of research worldwide. ReSA leads a number of task forces that assist in achieving these aims.

II. PROCESS

The FAIR4RS WG is enabling coordination and leverage existing community-led discussions on how to define and effectively apply FAIR principles to research software to achieve adoption of these principles. The FAIR4RS WG formed in 2019, and is on-track to deliver three outputs:

- A document developed with community support defining FAIR principles for research software
- A document providing guidelines on how to apply the FAIR principles for research software (based on existing frameworks)
- A document summarising the definition of the FAIR principles for research software, implementation guidelines and adoption examples.

The first of these outputs is expected to be completed by the third quarter of 2021 (it is currently going through a final, formal community review process [5]), and work is now commencing on the second. The resulting adoption and implementation of FAIR principles for research software will create significant outcomes for many stakeholders, including increased research reproducibility for research organizations, clarity for funders around their own requirements for software investments, and guidelines for publishers on sharing requirements. The FAIR4RS WG is recognised in the European Open Science Cloud (EOSC) FAIR Working Group's Six Recommendations for Implementation of FAIR Practice (2020) [4] as the primary community forum for this work.

To get to this point, the group leaders began by defining four topics that could be worked on independently. Each of these subgroups was led by a working group steering committee member, then advertised to the group members to find out who was interested. The groups met over about 9 months, with somewhat different means of working and communicating, but generally either used email or online chat to communicate, and mostly worked via shared documents, with tracked changes and comments in these documents also used for fine-grained communication. These subgroups and their outputs are:

- A fresh look at FAIR for Research Software. This subgroup examined the FAIR principles in the context of research software from scratch, not based on pre-existing work [2]. It published a report on its findings [3].
- FAIR work in other contexts This subgroup analyzed how FAIR principles are applied to research objects other than data/software. It's findings are captured in an unpublished final report.
- Research software definition. This subgroup reviewed existing definitions to specify the scope for the working group outputs. It is in the process of publishing a final report, and currently has produced a draft report, which is in the process of being published.
- New research related to FAIR Software. This group reviewed recent research and studies around FAIR software, via up-to-date identification of approaches that can help structure FAIR4RS work. It has produced a Zotero reading list and a draft report on important insights from the review and a survey, which is in the process of being published.

III. Community engagement

As the FAIR4RS WG nears completion of definition of nearly 12 months of work to develop community-endorsed principles for FAIR for Research Software, it is useful to reflect on how community engagement has occurred. This work requires engagement with many communities and across diverse geographic and domain backgrounds. The Fostering Fair Data Practices in Europe (FAIRsFAIR) report [6] on FAIRness of software strongly reinforces the importance of gaining community agreed application of the FAIR principles to research software, and its third recommendation provides guidance on how this should occur.

The FAIR4RS WG emerged from the "FAIR for research software" session at RDA Virtual Plenary 15 in April 2020, which was convened by the RDA Software Source Code Interest Group and with authors of the position paper "Towards FAIR Principles for Research Software" [2]. This group brought together and built on a wide range of work internationally. At the conclusion of this session, a Steering Committee was established of nine volunteers, who represent a wide range of research software interests, five of whom chose to be named as co-chairs of the group, but all of whom collaboratively co-lead the overall group. When one of the original steering committee members stepped down, the remaining members added a new person to bring the committee back to nine and to ensure good representation of the community.

The FAIR4RS WG has successfully built a global, interdisciplinary community that includes over 205 FAIR4RS WG members, plus over 90 individuals have attended and/or contributed to FAIR4RS discussions. The FAIR4RS WG aims to enable the participation of range of stakeholders, including those from different disciplines and geographic backgrounds:

1) Users of research software (external stakeholders)
2) Developers and/or maintainers of research software (internal stakeholders)
3) Creators/implementers of policy around research software and the personnel supporting it, and/or other research outputs for a team/department/organization/nation/region (e.g., managers of RSE teams, libraries, faculties) (internal stakeholders)
4) Managers of infrastructure that supports usage and/or development of research software and/or other research outputs (e.g., publishers, archives, repositories, registries, indices) (internal stakeholders)
5) Funders of research software and/or other research outputs (external stakeholders)

The FAIR4RS WG provides a range of ways for community members to engage according to their time and interests. The engagement approach utilised by the FAIR4RS WG is based on the Center for Scientific Collaboration and Community Engagement (CSCCE) participation model [7], which identifies four modes of stakeholder engagement, and the types of activities that may be relevant to each mode: convey/consume, contribute, collaborate and co-create. This has been adapted here to show the community activities that the FAIR4RS WG is engaging in, as shown in Fig. 1.

These activities are tracked on a page on the FAIR4RS WG GitHub repository. A key part of the work has been the process by which the FAIR4RS WG steering committee and WG members have worked together in multiple subgroups to analyse the state of the art and identify FAIR needs corresponding to the research software case. The FAIR4RS WG members have also undertaken nearly 40 public engagement activities over 15 months to publicise and/or gain feedback on this work. These are discussed in advance in monthly steering committee meetings (which have been public since 2021). When an opportunity is discussed, if someone chooses to lead it, others typically join the organization and presentation. This process has led to all of the steering committee members plus active non-steering-committee group members being involved in a variety of co-led engagement opportunities.

IV. Continued process

We've used engagement with the community during the process of writing the document defining the FAIR4RS principles as a mechanism for getting early community buy-in. This was supported by a number of key roles that were held by members of the FAIR4RS WG. A community manager has worked to ensure that the wider community is aware of FAIR4RS and the mechanisms for engaging with it. The subgroup leaders facilitated discussion of specific topics requiring further debate and inquiry, resulting in the production of key definitions and recommendations, as discussed above. Document editors applied these recommendations to the main documents, identified areas requiring wider community feedback, and responded to community comments to encourage discussion to enable consensus to be reached. In total, three rounds of community consultation took place. The first examined areas of disagreement or lack of consensus identified when the outputs of the subgroups were brought together and compared [8]. The second presented the community with different alternatives for

Convey/consume	Contribute	Collaborate	Co-create

Information can be accessed via:
- FAIR4RS WG email list via RDA posts
- #FAIR4RS on Twitter
- FAIR4RS WG website via RDA
- FAIR4RS Zenodo community
- Webinars
- Newsletters from stakeholders
- FAIR4RS Github
- FAIR4RS WG Steering Committee minutes

Input can be provided through:
- Town halls
- Surveys
- Google docs
- Gitter

Participation can be achieved through:
- Gitter
- Workshops
- Google docs
- Blog & article writing
- WG meetings

Joint leadership is available via:
- FAIR4RS Steering Committee membership
- Co-leading events
- Co-creating blogs & articles

Fig. 1: FAIR4RS working group community engagement activities

the scope, intent and wording of the draft FAIR4RS Principles, asking them to comment on which choice they felt was most appropriate. The third consultation, now underway, is a formal review by the various stakeholder communities of the full draft of the principles [5]. This is intended to get approval on the wording of the principles and explanatory text before the next stages, developing guidance to support adoption and implementation.

During this process, some individuals from the community made larger and more significant contributions, and were encouraged to engage more formally, through invitations to join WG meetings and to contribute directly to sections of the documents. This is similar to engagement models in open source projects and as mentioned above, where participants can be encouraged to progress from interested person through various levels to co-leader.

As commonly happens in large groups, recognizing contributions and defining co-authorship rules is not an easy task, particularly when people from different communities with different practices for co-authorship order contribute at different stages of the process. All contributions, from joining the working group via signing up to the mailing list, to participating in subgroups, events and consultations, to editing and authoring the FAIR4RS documents are recognized and recorded. Contributions at each stage are tracked according to the level of contribution, broadly in three categories: leadership roles, significant contributions (either by importance to the final documents or by consistent useful contributions), and other useful contributions. Contributors' placement in the author list are based on the level and amount of contributions they have made to the outputs at that point. The FAIR4RS WG is also listed as the last co-author, to recognise that all members play a part in ensuring that outputs are appropriate and adopted. This model ensures that all contributors are recognized and rewarded, and taking into account the challenges of managing a large number of individual contributors

V. NEXT STEPS

FAIR4RS WG will continue to use wide-scale community engagement to complete the rest of its planned outputs, on use cases and adoption guidelines. After gathering use cases and reviewing challenges regarding the implementation of the principles, the Steering Committee will produce a draft plan, which will be circulated and validated with the community

for feedback. At the same time, FAIR4RS will engage the community to get input on guidelines for application of definition of FAIR for research software. After gathering adoption examples from the community, the Steering Committee will draft a set of adoption guidelines, which will be circulated and validated with the community. The adoption plan will address how to work with different stakeholders, including those that will:

- Endorse and promote the guidelines
- Provide training on the guidelines
- Use the guidelines

It is anticipated that subgroups will be formed for these activities, as this structure has been effective to date. One useful change would be to widen the geographic diversity of participants. FAIR4RS's current 205 members come from 31 countries. However, 70% of them come from just six developed countries: Germany, United Kingdom, United States, France, The Netherlands and Australia). To assist in addressing this, FAIR4RS WG is now starting to schedule some events in French and Spanish. This aims to expand the group and interest in Africa and Latin America.

REFERENCES

[1] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne *et al.*, "The FAIR guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, 2016. doi: 10.1038/sdata.2016.18

[2] A.-L. Lamprecht, L. Garcia, M. Kuzak, C. Martinez, R. Arcila, E. M. del Pico, V. Angel, S. van de Sandt, J. Ison, P. Martinez, P. McQuilton, A. Valencia, J. Harrow, F. Psomopoulos, J. Gelpi, N. P. C. Hong, C. Goble, and S. Capella-Gutiérrez, "Towards FAIR principles for

research software," *Data Science*, vol. 3, pp. 37–59, 2020. doi: 10.3233/DS-190026

[3] D. S. Katz, M. Gruenpeter, T. Honeyman, L. J. Hwang, M. D. Wilkinson, V. Sochat, H. Anzt, and C. A. Goble, "A fresh look at FAIR for research software," *arXiv*, vol. abs/2101.10883, Jan. 2021.

[4] Directorate-General for Research and Innovation (European Commission) and EOSC Executive Board, "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 of the EU*, Oct. 2020. doi: 10.2777/986252

[5] N. P. C. Hong, D. S. Katz, M. Barker, A.-L. Lamprecht, C. Martinez, F. E. Psomopoulos, J. Harrow, L. J. Castro, M. Gruenpeter, P. A. Martinez, T. Honeyman *et al.*, "FAIR principles for research software (FAIR4RS principles)," *Research Data Alliance*, 2021. doi: 10.15497/RDA00065

[6] M. Gruenpeter, R. Di Cosmo, H. Koers, P. Herterich, R. Hooft, J. Parland-von Essen, J. Tana, T. Aalto, and S. Jones, "M2.15 assessment report on 'FAIRness of software'," *Zenodo*, Oct. 2020. doi: 10.5281/zenodo.4095092

[7] L. Woodley and K. Pratt, "The CSCCE Community Participation Model – A framework to describe member engagement and information flow in STEM communities," *Zenodo*, Aug. 2020. doi: 10.5281/zenodo.3997802

[8] D. S. Katz, N. P. Chue Hong, M. Barker, and M. Gruenpeter, "FAIR4RS WG subgroup community consultation March 2021," *Zenodo*, Mar. 2021. doi: 10.5281/zenodo.4635410 Specify grants which have funded this research other than those listed in grants: Alfred P. Sloan Foundation grant 2021-14116 Wellcome Trust grant 222436/Z/21/Z.