

## FAIRsFAIR<sup>1</sup> Project Comments on Data Repository Selection-Criteria That Matter<sup>2</sup>

The FAIRsFAIR project would like to thank the authors for the opportunity to respond to this updated proposal. Feedback to this version is an update of the text previously shared with the authors<sup>3</sup>

Criteria, characteristics and requirements related to data objects, data services, including repositories, are at the heart of the FAIRsFAIR project, and we welcome this proposal from important stakeholders in the research data lifecycle. We note that though the authors state that their “ultimate aim is to enable FAIR data” and “many repositories are beginning to incorporate the FAIR Principles”, these criteria do not address any specific alignment with the FAIR principles.

Below we provide a general response to the paper, identify alignment with ongoing FAIRsFAIR work and make some broad comments on the proposed criteria. Based on the authorship and content of the criteria we assume there are two key areas for further work and clarification.

**First, the role of FAIRsharing as a registry and how these repository characteristics can be integrated into a schema for common repository metadata declaration, validation and management.** This work is critical to the availability of machine-actionable information related to repositories and their collections as well as being a dependency for user experience design (including through filters etc.).

This repository-focussed perspective maps closely to FAIRsFAIR work to define assessment criteria for services and repositories. FAIRsFAIR will provide an improved description of data repositories, additional components and properties to the metadata schema, and tools that enable reuse to align with FAIR data practices. The co-dependent area of work in FAIR/EOSCs the assessment and validation of digital object characteristics (ultimately including both research data and published papers as linked entities). FAIRsFAIR would value further information on how the authors envisage the long-term validation and management of common repository and object characteristics that meet registry and publisher needs, as well as those of data users.

**The second area of focus is to identify the needs of key stakeholders.** The Turning FAIR into Reality report<sup>4</sup> explicitly includes the publishers (of data and publications), alongside others, for example institutions and funders. It would be greatly appreciated if the publishers among the authors could share an idealised ‘workflow’ description of how key actors (authors, publishers, funders, institutions, curators, data stewards) and entities (papers, research data, repository metadata) would interact to meet their needs. This can

---

<sup>1</sup> <https://www.fairsfair.eu/>

<sup>2</sup> <https://osf.io/m2bce/>

<sup>3</sup> <https://doi.org/10.5281/zenodo.3632985>

<sup>4</sup> [Turning FAIR into reality - Publications Office of the EU](#)

then be integrated as a use case into the iterative development of metadata, assessment criteria, processes and evidence within FAIRsFAIR and elsewhere.

The project also has an excellent opportunity to promote and develop these publisher criteria, alongside those required by other stakeholders through its repository support activities.

The paper implies that the outcomes from the author perspective will be a set of metadata and supporting filter options for FAIRsharing, which align with the common criteria agreed by publishers. If this is correct we and others will be keen to follow progress, is there a planned communications channel? We would value an opportunity to engage with this process. A clear timeline that included a target date for a stable set of proposed criteria would allow us to integrate these into our work on the requirements for FAIR objects and FAIR enabling data services, including repositories.

We note that this version of the text no longer states (e.g. in relation to the presence or absence of certification) that “it is not yet practical to filter on some criteria because they are not widely used by data repositories”. The text “at this stage, this criteria is for information only” has been inserted but it is not clear whether these fields would be excluded from FAIRsharing filters or simply excluded from Publisher evaluations of whether a repository should be ‘recommended’. The paper acknowledges that there will not be a single list of recommended repositories (“different publishers may attribute more importance to some criteria than others”) but it would be interesting to understand how the user experience within FAIRsharing is expected to work. For example how an institutional repository (specifically excluded from scope by the authors) that demonstrated all of the required criteria would be excluded from filter results. It would also be interesting to see further reflection on future challenges for FAIRsharing in meeting the needs of institutions and funders as users of their filtering mechanisms, considering these stakeholders may also attribute more importance to some criteria than others.

The new separation of criteria is not always clear, for instance the presence of a persistent identifier is under ‘Repository-level Criteria’ while a repository process to support publisher needs (Data Access for Pre-Publication Review) is listed under Dataset-level Criteria.

Only ‘Data Access for Pre-Publication Review’ is strictly limited to the Publisher workflow context. The latest version helpfully clarifies that the “user” in ‘User Support’ applies to both data depositors and data users. .

‘FAIRsharing Record Maintainer’ might be a useful starting point to consider the range of roles needed to be sure that registry information is appropriately collected, maintained, versioned, and used. This is important if values are the result of self-declaration (repository status, data curation) or consist of yes/no responses (preservation plan).

But other than these examples the criteria are of value beyond the publisher/registry workflows to other stakeholders which should not be ignored and are already present in a range of recommendations and standards for repositories and data management e.g. persistent identification, data and metadata community standards, rights information, preservation policies, resource sustainability, funding.

The one new item we have noted in this version of the criteria is 'Repository coverage' ("The higher-level subject areas/disciplines the repository covers, as well as cross-disciplinary domains, such as the types or data, technology and study"). The authors will no doubt recognise the wide implications of this addition, in terms of both real world research infrastructure and the development of a suitable metadata and ontology approach. FAIRsFAIR would value the opportunity to collaborate on defining those implications and developing a broadly applicable approach. The criteria could usefully be mapped to the current FAIRsharing schema for 'databases'<sup>5</sup> as a starting point for formal design and management.

This identification of so many aligned criteria represents a valuable outcome of the activities by the authors of the paper and their wider collaborators. We would urge them to consider developing their work through feedback directly to the standards they align closely with including the CoreTrustSeal and through the collaborative development of appropriate metadata elements and controlled vocabularies to support publishers and the wider stakeholders in research infrastructure.

---

<sup>5</sup> <https://fairsharing.org/databases/>