# Repository Features to Help Researchers: An invitation to a dialogue

**Publishers** (repocriteria@googlegroups.com): Matthew Cannon1 (0000-0002-1496-8392), Chris Graf2 (0000-0002-4699-4333), Kiera McNeice3 (0000-0003-2839-4067), Wei Mun Chan4 (0000-0002-9971-813X), Sarah Callaghan5 (0000-0002-0517-1031), Ilaria Carnevale5 (0000-0001-8509-0495), Imogen Cranston6 (0000-0002-7134-499X), Scott C. Edmunds7 (0000-0001-6444-1436), Nicholas Everitt1 (0000-0001-8343-8910), Emma Ganley8,9 (0000-0002-2557-6204), Iain Hrynaszkiewicz8 (0000-0002-9673-5559), Varsha K. Khodiyar10 (0000-0002-2743-6918), Adam Leary11 (0000-0002-9147-8397), Thomas Lemberger12 (0000-0002-2499-4025), Catriona J. MacCallum13 (0000-0001-9623-2225), Hollydawn Murray6,14 (0000-0002-8243-2493), Kathryn Sharples2 (0000-0003-2809-6828), Marina Soares E Silva5 (0000-0001-9530-627X), Guillaume Wright6  (0000-0003-0166-9200).

**Moderators**: Susanna-Assunta Sansone15 (0000-0001-5306-5690), Peter McQuilton15 (0000-0003-2687-1982)

1. Taylor & Francis, Park Square, Milton Park, Abingdon, OX14 4RN, UK.
2. Wiley, 9600 Garsington Road, Oxford, OX4 2DQ, UK.
3. Cambridge University Press, Shaftesbury Rd, Cambridge, CB2 8BS, UK.
4. eLife Sciences Publications, Ltd, Westbrook Centre, Milton Road, Cambridge, CB4 1YG, UK.
5. Elsevier, Radarweg 29, 1043NX, Amsterdam, The Netherlands.
6. F1000 Research Ltd, 5 Howick Place, London, SW1P 1WG.
7. GigaScience, BGI Hong Kong Tech Co Ltd., 26F A Kings Wing Plaza, 1 On Kwan St, Shek Mun, N.T., Hong Kong, China.
8. PLOS (Public Library of Science), Carlyle House, Carlyle Road, Cambridge CB4 3DN, UK.
9. (present address) protocols.io, 2120 University Ave, Suite 625, Berkeley, CA, USA.
10. Springer Nature, 4 Crinan Street, London, N1 9XW, UK.
11. Oxford University Press, Great Clarendon Street, Oxford, OX2 6DP, UK.
12. EMBO Press, Meyerhofstrasse 1, 69117 Heidelberg, Germany.
13. Hindawi Ltd, 1 Fitzroy Square, London, W1T 5HF, UK.
14. (present address) HDR UK
15. Oxford e-Research Centre, Department of Engineering Science, University of Oxford, Oxford, OX1 3QG, UK

## 

## **Background and Motivation**

The data repository landscape is broad and complex, with several thousand repositories currently in existence. These range from generalist to discipline- or data-specific; from open to restricted; from non-profit to commercial; and they may be run by academic institutions, governments, companies, or others. Some data repositories are open to deposition from all researchers with relevant datasets, whilst others will only accept data from researchers affiliated with a specific institution, project, funder or from a particular nation. To support the Findability, Accessibility, Interoperability and Reusability of data, many data repositories are beginning to incorporate the FAIR Principles[[1]](#footnote-1) in their policies and implement the necessary technical enhancements.

However, we are acutely aware of the inequity this has the potential to create, especially in low- and middle-income countries, unless there is appropriate support from all stakeholders at national, international and global levels to implement enhancements to comply with the FAIR principles. These issues must be addressed in tandem as standards such as FAIR become more accepted in for example Western Europe and the US, where most of the organisations co-authoring this article are based. As we transition to a global and interoperable infrastructure for open research, it is important that appropriate support and resources are given to those who are beginning to develop policies and a roadmap for their own context. All data repositories play a key role in ensuring greater transparency and preservation of the information that underpins research findings – goals that are of great importance to funders, governments, and all proponents of good research practice.

Although there are several factors researchers may need to consider when choosing how to share their data, including institutional or funder policies, publishers are also regularly approached by researchers asking for advice about which repository they should use for their data[[2]](#footnote-2). As publishers are committed to playing their part in the evolution of a more open research landscape, many are developing policies that encourage researchers to make data and other digital materials accessible via appropriate repositories, and encourage the formal declaration of these scholarly outputs in articles[[3]](#footnote-3). However, given the complexity of the data repository landscape, the question of what constitutes an “appropriate repository” for a given set of digital materials is often non-trivial.

To date publishers have adopted a variety of approaches to supporting researchers in this context. Some have maintained manually-curated lists of repositories appropriate to specific journals or disciplines. Sometimes journals have incorporated into policy the recommendations of their editorial boards about appropriate repositories for particular domains or data types.[[4]](#footnote-4) Others maintain more general guidance about considerations authors may have when choosing a repository, including existing initiatives to assess and certify the quality of repositories. However with thousands of repository options available, and variation among the communities each publisher serves, the result has been a fragmented landscape of inconsistent publisher advice[[5]](#footnote-5) provided to authors, creating additional complexity and confusion for researchers.

Recognising this, the publishers and authors of this paper came together to discuss how we could reduce complexity and inconsistency in publisher advice to better support researchers in selecting an appropriate data repository. It is a shared goal among publishers and other stakeholders to increase repository use – which remains far from optimal – and we assume that helping researchers find a suitable repository more easily will help achieve this. This harmonisation is critical and timely, particularly during the unfolding health emergency of the COVID-19 pandemic, which has seen increased demand from funders and publishers for rapid results and sharing of datae.g.,[[6]](#footnote-6),[[7]](#footnote-7),[[8]](#footnote-8). Expedited data-sharing should not however happen at the expense of reusability, and the key to ensuring data quality control and sustained access is the deposition of data in appropriate repositories.

**The role of the RDA FAIRsharing WG**

The impetus for this work began in 2017 with a discussion between publishers and members of the FORCE11 Data Citation Implementation Pilot (DCIP) group[[9]](#footnote-9),[[10]](#footnote-10) , and of the joint FORCE11 and RDA FAIRsharing Working Group (WG)[[11]](#footnote-11) on the question of developing a shared list of recommended data deposition repositories. The work commenced in January 2018, and representatives of the Force11/RDA FAIRsharing WG’s chairs acted as neutral coordinators, facilitating discussions to find common ground, and helping reaching out to other stakeholders.

The Force11/RDA FAIRsharing WG has operated under RDA since 2015, the group was set up to foster a culture change where the use of standards, databases and repositories is pervasive and seamless. Matured under the RDA, the FAIRsharing registry[[12]](#footnote-12) is the key output of the Force11/RDA FAIRsharing WG. FAIRsharing describes data and metadata standards (for citation and reporting), repositories (databases and knowledgebases) and data policies, and captures their interrelations to: enable (i) consumers to discover, select and use these resources with confidence, and (ii) the producers of these resources to make them more discoverable, adopted and cited. Nowadays FAIRsharing is recommended and used by many stakeholders in data management and scholarly communication[[13]](#footnote-13).

From the RDA FAIRsharing WG point of view, the ultimate objective is to map repository features across initiatives, and to identify a common core set of metadata fields that all stakeholders want to see in the FAIRsharing registry. The FAIRsharing registry is agnostic as to the selection process of standards, repositories, and policies. As part of its commitment to working with and for all stakeholder groups, FAIRsharing will implement these common metadata fields progressively to help users, including researchers and the people who advise them, to better understand the features of each repository, and assist repository maintainers to evaluate and improve (where necessary) their current operations.

**Community Engagement, Approach and Timeline**

A key concern from the beginning of this work was to be mindful and supportive of existing and new initiatives, some of which emerged during the period that this discussion among publishers was progressing. When the work started the only existing published guidelines were criteria to recommend and evaluate the trustworthiness of data repositories[[14]](#footnote-14), and the domain-specific ELIXIR Deposition Repositories, published in 2017[[15]](#footnote-15). Subsequently, the domain-agnostic CoreTrustSeal (CTS) was set an as entity in 2018[[16]](#footnote-16), and in 2020 the TRUST principles were published[[17]](#footnote-17), along with the COAR guidelines[[18]](#footnote-18); in 2021 the funder-specific guidance from the USA National Institutes of Health were released[[19]](#footnote-19), along with the Science Europe criteria for selection of trustworthy repositories[[20]](#footnote-20).

Such diversity of parallel efforts shows that this issue is important to a number of stakeholders who have been working within their own communities, and tackling the challenge from their own viewpoints. In many cases publishers also highlighted these efforts as part of their existing guidance for researchers on repository selection. Additionally, publishers and authors of these papers have also contributed to some of these nascent initiatives, in particular the TRUST principles and the Science Europe guidelines.

Universal standards that would be appropriate to all researchers in all situations have yet to emerge from these multiple initiatives, and so there remains a need for greater consistency and transparency in the recommendations publishers make regarding data repositories. A key goal of this work was therefore to identify and agree on the features publishers recommend considering when advising authors on how best to select a data repository, within the context of the current data repository landscape.

As described above, in 2017 publisher members of the FORCE11 DCIP group, and of the joint FORCE11 and RDA FAIRsharing WG began discussing a shared list of recommended data deposition repositories, which led to the first phase of this work focused on defining our objectives and relations with other efforts, and on teasing out an initial list of features from the needs and views of the represented journals and publishers. This list was presented at the 12th and 13th RDA Plenaries (2018 and 2019) and shared internally within the participating organisations. Concurrently, an article[[21]](#footnote-21) authored by the over 60 FAIRsharing adopters (including funders, repositories, and standards developers, as well as research infrastructures and publishers) delivered an analysis of the data policies of major data-focused journals and publishers operating in the life sciences, and the repositories and standards they recommend. The results confirmed discrepancies in resource recommendations across the data policies.

The draft list of features was reviewed and refined iteratively by the kickoff group. This process was further refined through a collaboration[[22]](#footnote-22) between FAIRsharing and DataCite and the participation of more publisher representatives, all of whom are actively implementing data policies and recommending data repositories to researchers.

In October 2019, the group published the work as v1 and v2[[23]](#footnote-23) and presented it at the 14th RDA Plenary, in a joint session between the FAIRsharing WG, the RDA/WDS Certification of Digital Repositories IG, and the Data Policy Standardisation and Implementation IG. In November 2019, a survey was organized to explicitly solicit community comments. The survey was disseminated via RDA and other email lists, as well as via social media channels and blog posts by DataCite[[24]](#footnote-24) and the publishers[[25]](#footnote-25). This resulted in 52 responses, 67.9% of which were repository maintainers. Others were individuals and relevant organisations such as CTS and ELIXIR. In February 2020, 120 comments were extracted from the responses, classified, and triaged. The vast majority of responses were positive, suggesting clarifications and improvements to the presentation of the activity and the list of features. These comments were addressed in a revised preprint published in October 2020[[26]](#footnote-26). In November 2020, the group presented this work at the 16th RDA Plenary[[27]](#footnote-27).

Based on further feedback and concerns raised by the community in late 2020 and early 2021, notably from COAR[[28]](#footnote-28) (signed by other organisations such as LIBSENSE and Science Europe) and the FAIRsFAIR project[[29]](#footnote-29), the publishers have published a response[[30]](#footnote-30) to clarify their motivation and this preprint has been further revised to clarify the motivations and goals of publisher representatives in this work.

## **Core Repository Features**

The below table outlines a list of features that should be considered by researchers when choosing a data repository. For each feature the table outlines:

* A brief definition of the feature, and what information or functionality it provides.
* Values to describe the status of each feature for a given repository. For example, the possible values for a feature might be “Yes” or “No” to indicate the presence or absence of that feature.

This list of features is intended as a framework within which publishers can make *recommendations* to researchers, not as a way to restrict which repositories researchers may choose for their data. This feature list also serves as a framework for registries of repositories, such as the FAIRsharing registry, to organize and index repositories according to a core set of attributes that are relevant to both researchers and publishers.

Generally, by agreeing on a finite list of features for consideration, publishers can reduce inconsistencies and complexity when developing data sharing policies to support open research. The aim especially is to raise awareness of data stewardship to researchers not yet familiar with data repositories or policies around data sharing and to provide some consistency in the guidelines they are likely to encounter among publishers. A common, openly-available list also enables publishers to be more transparent about which features are considered when making recommendations. Our ultimate aim is to enable FAIR data, and standardising how publishers recommend repositories will reinforce the use and value of repositories and bring us closer to this goal.

### **Addressing community concerns**

The authors acknowledge that concerns have been raised by the community about the scope and specificity of this list of features. We recognise that in many cases while the listed features are desirable, their adoption is not yet widespread enough for it to be sensible or practical to recommend only those repositories which have implemented them.

Furthermore, in most cases the agreed “ideal” value for filtering is simply the presence of information. For example, the ideal value we recommend for Data Access Condition is “Yes” to indicate the *presence of clearly stated information* about data access conditions and processes, rather than any expectations regarding what the actual conditions or processes are. We believe that having transparent information about what those conditions are will help researchers make an informed decision.

The features we highlight here are very much those that capture the concerns of publishers trying to implement data sharing policies whilst providing guidance to researchers in a very complex landscape. Our intention is that the features we highlight will act to initiate engagement and collaboration among publishers, repositories and the RPOs, government and funders that ultimately make the policies around Open Research. The list of features is intended to start a dialogue, and is not proposed as a set of criteria[[31]](#footnote-31) to be immediately implemented. As we start this conversation, it is important that we act together with other stakeholders to raise awareness of the challenges involved around FAIR data and to prevent any perverse consequences. The features we have highlighted are those that some publishers believe are important for repositories to make transparent, enabling researchers to consider aspects such as domain relevance, funder, and institutional requirements, in order to select the most suitable data repository for their needs. This does not mean that repositories without these features should not be supported. Our aim is to highlight those features that are most relevant to publishers developing and implementing data sharing policies to support open research, and that map onto existing guidelines elsewhere.

As appropriate, features have also been mapped to the relevant TRUST principles, the CTS requirements, the COAR guidelines, the NIH and the ELIXIR feature lists (see [Mapping file](https://docs.google.com/spreadsheets/d/19bptRYOBdlsEg6Hwv7m2gO1Z1EHJPeLYan1AuRRUmNA/edit#gid=0)). Many of our proposed features match the existing framework, except for two features: “Data Access for Pre-Publication Review” and “Data Deposition Condition”.

|  |  |  |
| --- | --- | --- |
| **FEATURE** | **DEFINITION** | **VALUES IN REPOSITORY INDEXING METADATA FIELD** |
| **Repository-level** | | |
| **Certification and Community Badge[[32]](#footnote-32),[[33]](#footnote-33)** | Certification schemes and/or community badges that assess certain aspects of the repository (e.g., its fitness, trustworthiness, adoption):  Does the repository have a certification or community badge? | - Presence ‘yes’, or absence ‘no’.  - Type of certification or badge.  - URL to the certification or badge. |
| **Data Access Condition[[34]](#footnote-34),[[35]](#footnote-35)** | Data access mechanisms and terms to define access at repository and/or dataset level.  Does the repository explicitly state data access conditions on dataset landing pages? | - Presence ‘yes’, or absence ‘no’.  - Types of data access conditions the repository is able to facilitate, e.g. ‘open’, ‘controlled’  - URL to the condition information. |
| **Data Access for Pre-Publication Review** | Mechanism and process for sharing deposited data via a link anonymously (or otherwise depending on journal policy regarding open/closed review).  Does the repository have a mechanism to facilitate peer review of embargoed data? | - Presence ‘yes’, or absence ‘no’.  - Type of mechanism (e.g. private reviewer URLs). |
| **Data and Metadata Standards[[36]](#footnote-36),[[37]](#footnote-37),[[38]](#footnote-38)** | The community-defined standards the repository implements to enable the representation of data and/or metadata in a consistent, machine readable form (e.g. via models, formats, schemas, vocabularies, ontologies). These standards facilitate the discovery and interpretation of data and/or metadata.  Which data and metadata standards (if any) has the repository implemented? | - Presence ‘yes’, or absence ‘no’.  - Type of standards.  - Relations between types of standards, and between standards and repositories (to track their use and adoption). |
| **Data Curation[[39]](#footnote-39),[[40]](#footnote-40),[[41]](#footnote-41),[[42]](#footnote-42)** | Review and annotation of the data performed by the repository (e.g. via a data submission tool that enforces some curation, or by its curation team).  Does the repository curate its holdings? | - Presence ‘yes’, or absence ‘no’.  - URL to the curation information. |
| **Funding[[43]](#footnote-43)** | The type of funding (e.g. grants, donations, memberships) and the organisation(s) that fund the repository.  Who funds the operation of the repository? | - Types of funding  - Name of the funding organization (e.g. using FundRef) |
| **Persistent Identifiers for Data[[44]](#footnote-44),[[45]](#footnote-45)** | Globally unique and Persistent IDentifiers (PIDs).  Does the repository assign PIDs to it’s holdings? If so, which PID schema has been implemented? | - Presence ‘yes’, or absence ‘no’.  - Type of PID schema. |
| **Repository Coverage[[46]](#footnote-46),[[47]](#footnote-47)** | The higher-level subject areas/disciplines the repository covers, as well as cross-disciplinary domains, such as the types or data, technology and study.  What is the disciplinary scope of the repository? | - Type of disciplines  - Type of domains  - Taxonomic range, if applicable |
| **Repository Status[[48]](#footnote-48)** | The life cycle status of the repository Repositories may be undergoing (re)development, enhancement or maintenance, as happens in any mature system.  Is the repository currently accepting data submissions? Is the repository under development? Has the repository been deprecated? | -‘Ready’ (production-level).  - ‘In development’ (actively being developed but still not ready for use).  - ‘Deprecated’ (no longer maintained or superseded).  - ‘Uncertain’ (when the life cycle position is unclear). |
| **Resource Sustainability[[49]](#footnote-49),[[50]](#footnote-50)** | Plan that gives information about sustainability plans for the repository.  Does the repository provide information on it’s sustainability plans? | - Presence ‘yes’, or absence ‘no’. |
| **User Support[[51]](#footnote-51)** | Support to users during or after submission.  Does the repository have a contact point (e.g. helpdesk email or contact form) to assist data depositors and data users? | - Presence ‘yes’, or absence ‘no’. |
| **Record Maintainer[[52]](#footnote-52)** | Contact (person or organization) for the record in a registry that describes the repository. The participation of the owner or maintainer of the repository helps verify the information in the registry.  Has the owner or maintainer of the repository claimed the registry record? | - Presence ‘yes’, or absence ‘no’.  - Contact details (incl. ORCID). |
| **Dataset-level** | | |
| **Citation to related publications** | A mechanism to link datasets to related articles or pre-prints.  Does the repository enable data to article linking? At what stage of data deposition is article information required? | - Presence ‘yes’, or absence ‘no’.  - Type of mechanism. |
| **Data Contact Information** | Contact info (for the person, depositor, producer or owner, ideally with ORCID; or organization) of the data.  Does the repository show data depositor contact information on dataset landing pages? | - Presence ‘yes’, or absence ‘no’. |
| **Data Deposition Condition** | Deposition of data.  Who can deposit data into the repository? | - Presence ‘yes’, or absence ‘no’.  - Type of condition.  - URL to the condition information. |
| **Data Preservation Policy[[53]](#footnote-53),[[54]](#footnote-54),[[55]](#footnote-55)** | Policy that details how the preservation of the data is ensured.  Does the repository provide information on its data preservation policies? | - Presence ‘yes’, or absence ‘no’.  - URL to the policy. |
| **Data Reuse Condition[[56]](#footnote-56),[[57]](#footnote-57)** | Licence or terms of use for reuse of existing data in the repository; these can be the same for every dataset in the repository, or vary from dataset to dataset.  What are the range of data reuse conditions which may be applied to the data holdings? | - Presence ‘yes’, or absence ‘no’.  - Type of condition.  - URL to the condition information. |
| **Data Versioning[[58]](#footnote-58)** | Mechanism and process to make and track edits to a dataset after deposition.  Does the repository enable modifications to published data (e.g., to correct it or append additional information)? Is there a process to distinguish, link and access all public versions of the data? | - Presence ‘yes’, or absence ‘no’. |

## **Challenges and Future Work**

This work and the feedback we have received from the community have highlighted a number of challenges that must be addressed, but we consider out of scope for this work:

* *Licence equivalence:* many repositories do not use Creative Commons (CC) licences but state they either use something ‘equivalent’, or have custom licences. However, this equivalence is difficult to measure and is something we need to be able to represent to allow comparison between different terms and conditions. Standard licences and comprehensive licence mapping would help reduce the number of licences data users need to deal with, and overall aid with interoperability and automation issues.
* *File size and cost associated with deposition:* this is key information for journals and publishers, as well as for funders and researchers, but is hard to collect as often this information is not publicly available, or may differ for different repository users (e.g., academic versus commercial rates and special deals).
* *Institutional repositories:* the authors believe that institutional repositories should be considered and assessed in the same way as any other repository by researchers looking for an appropriate place to deposit data. However, ensuring equal visibility and relevance of institutional repositories to a publisher’s international audience of researchers may be difficult in practice.
* *Data/dataset citation:* whether and how a deposition repository follows the data citation principles and has sufficient metadata in each dataset’s landing page are other ideal features, but currently this information is not easily collected.
* *Data usage:* metrics to assess data use in the repository to allow researchers insight into data reuse. Whether it is feasible or technically possible for a repository to provide this information.

## **Next Steps**

Although the authors have sought and responded to feedback from the wider data-sharing community during the course of this work, we acknowledge that it still primarily represents the perspective of publishers, who are by no means the only stakeholders when it comes to recommending data repositories. While we did attempt to use our networks to solicit feedback we understand that some subject fields have been better represented than others, and we are keen to work to address this.

The main goal of the authors of this work remains consistency and transparency in the recommendations publishers make about repositories. We believe this goal will reduce complexity and confusion for researchers navigating the data-sharing landscape, leading to greater use of repositories and thereby FAIRer data. We recognise that this goal must be pursued in close collaboration with other stakeholders to ensure that our work benefits the data-sharing community as a whole. Therefore, we invite a wide, multi-stakeholder consultation to ensure that we can collectively arrive at recommendations that speak for and can be endorsed by the whole community, and ultimately will clarify and improve the guidance that is offered to researchers.

Our immediate next step will be to engage directly with the wider data-sharing community at a session in the upcoming 17th RDA plenary. The session[[59]](#footnote-59), organized by the RDA FAIRsharing WG, aims to map the landscape of these initiatives, bring them together, raise awareness, understand their scope and motivation, and see if and how they are connected or complementary. The RDA FAIRsharing WG email list will also be used to continue the discussion and we invite interested parties to join in[[60]](#footnote-60).

## **Acknowledgements**

We would like to thank DataCite, represented by Helena Cousijn, for their participation in and contributions to this project, David Carr from Wellcome who also participated in discussions.

1. <https://doi.org/10.1038/sdata.2016.18> [↑](#footnote-ref-1)
2. <https://doi.org/10.6084/m9.figshare.13227875.v2> [↑](#footnote-ref-2)
3. <http://doi.org/10.5334/dsj-2020-005> [↑](#footnote-ref-3)
4. See for example, the Joint Data Archiving Policy adopted by a number of journals in evolutionary biology and ecology in support of the Dryad Data Repository: <https://datadryad.org/docs/JointDataArchivingPolicy.pdf> [↑](#footnote-ref-4)
5. <https://fairsharing.org/recommendations/> [↑](#footnote-ref-5)
6. <https://wellcome.ac.uk/coronavirus-covid-19/open-data> [↑](#footnote-ref-6)
7. <https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-guidelines-oa-covid-19_en.pdf> [↑](#footnote-ref-7)
8. C19 Rapid Review Initiative: <https://oaspa.org/data-deposition-required-for-all-c19-rapid-review-publishers> [↑](#footnote-ref-8)
9. <https://www.force11.org/group/dcip> [↑](#footnote-ref-9)
10. <http://dx.doi.org/10.1038/s41597-019-0031-8> [↑](#footnote-ref-10)
11. <https://rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html> [↑](#footnote-ref-11)
12. <https://fairsharing.org> [↑](#footnote-ref-12)
13. <https://fairsharing.org/communities#adopters> [↑](#footnote-ref-13)
14. <https://doi.org/10.2218/ijdc.v9i1.309> [↑](#footnote-ref-14)
15. <https://elixir-europe.org/platforms/data/elixir-deposition-databases> [↑](#footnote-ref-15)
16. <https://www.coretrustseal.org> [↑](#footnote-ref-16)
17. <https://doi.org/10.1038/s41597-020-0486-7> [↑](#footnote-ref-17)
18. <https://doi.org/10.5281/zenodo.4110829> [↑](#footnote-ref-18)
19. <https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html> [↑](#footnote-ref-19)
20. <https://scienceeurope.org/our-priorities/research-data/research-data-management> [↑](#footnote-ref-20)
21. <https://doi.org/10.1038/s41587-019-0080-8> [↑](#footnote-ref-21)
22. https://doi.org/10.5438/z32p-wj46 [↑](#footnote-ref-22)
23. <https://osf.io/m2bce/?show=revision> [↑](#footnote-ref-23)
24. Datacite (<https://doi.org/10.5438/mbs5-tz08>) [↑](#footnote-ref-24)
25. Cambridge University Press (<https://www.cambridge.org/core/blog/2019/12/10/openresearch-selecting-a-data-repository-criteria-that-matter>); eLife (<https://elifesciences.org/inside-elife/7b9ba7ef/elife-latest-criteria-for-data-repository-selection>); Elsevier (<https://www.elsevier.com/connect/share-your-thoughts-to-make-data-sharing-simpler-and-more-efficient>); F1000Research (<https://blog.f1000.com/2019/11/29/data-repository-selection-criteria-that-matter-request-for-comments>); GigaScience (<http://gigasciencejournal.com/blog/fairsharing-data-repository-selection>); Hindawi (<https://about.hindawi.com/blog/data-repository-selection-criteria-that-matter-request-for-comments>); PLOS (<https://blogs.plos.org/plos/2019/11/request-for-comments-on-data-repository-selection-criteria-that-matter>); Springer Nature (<https://researchdata.springernature.com/users/69696-varsha-khodiyar/posts/57690-data-repository-selection-request-for-comments>); Taylor and Francis (<https://librarianresources.taylorandfrancis.com/data-repository-selection-criteria-that-matter>); Wiley (<https://www.wiley.com/network/researchers/latest-content/data-repository-selection-criteria-that-matter-we-want-to-hear-from-you>). [↑](#footnote-ref-25)
26. <https://doi.org/10.5281/zenodo.4084763> [↑](#footnote-ref-26)
27. In a Funders policy session led by the Research Funders and Stakeholders on Open Research and Data Management Policies and Practices IG. [↑](#footnote-ref-27)
28. <https://www.coar-repositories.org/news-updates/input-to-data-repository-selection-criteria-that-matter> [↑](#footnote-ref-28)
29. <https://doi.org/10.5281/zenodo.4630588> [↑](#footnote-ref-29)
30. <https://doi.org/10.5281/zenodo.4458082> [↑](#footnote-ref-30)
31. The authors recognise that the title of a previous version of the pre-print ‘Criteria that Matter’ raised concerns and obscured the intentions of the piece in starting a discussion on how to improve guidance for researchers. [↑](#footnote-ref-31)
32. TRUST principle - Responsibility: to be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service. [↑](#footnote-ref-32)
33. These include CTS, ELIXIR Deposition Repositories, and the funder-specific guidance such as those from the USA NIH. [↑](#footnote-ref-33)
34. CTS R2 requirement - Licence: the repository maintains all applicable licenses covering data access and use and monitors compliance. [↑](#footnote-ref-34)
35. CTS R4 requirement - Confidentiality/Ethics: the repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms. [↑](#footnote-ref-35)
36. CTS R7 requirement - Data integrity and authenticity: the repository guarantees the integrity and authenticity of the data. [↑](#footnote-ref-36)
37. CTS R8 requirement - Appraisal: the repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users. [↑](#footnote-ref-37)
38. CTS R14 requirement - Data reuse: the repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data. [↑](#footnote-ref-38)
39. TRUST principle - Responsibility: to be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service. [↑](#footnote-ref-39)
40. CTS R0 requirement - Context: the repository level of curation [↑](#footnote-ref-40)
41. CTS R8 requirement - Appraisal: the repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users. [↑](#footnote-ref-41)
42. CTS R11 requirement - Data quality: the repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality- related evaluations. [↑](#footnote-ref-42)
43. CTS R5 requirement - Organizational infrastructure: the repository has adequate funding (and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission). [↑](#footnote-ref-43)
44. TRUST principle - Responsibility: to be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service. [↑](#footnote-ref-44)
45. CTS R0 requirement - Data discovery and identification: the repository enables users to discover the data and refer to them in a persistent way through proper citation. [↑](#footnote-ref-45)
46. TRUST principle - Transparency: to be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence. [↑](#footnote-ref-46)
47. CTS R0 requirement - Context: the repository type [↑](#footnote-ref-47)
48. TRUST principle - Technology: to provide infrastructure and capabilities to support secure, persistent, and reliable services. [↑](#footnote-ref-48)
49. TRUST principle - Sustainability: to sustain services and preserve data holdings for the long-term. [↑](#footnote-ref-49)
50. CTS R3 requirement - Continuity of Access: the repository has a continuity plan to ensure ongoing access to and preservation of its holdings. [↑](#footnote-ref-50)
51. TRUST principle - User Focus: to ensure that the data management norms and expectations of target user communities are met. [↑](#footnote-ref-51)
52. TRUST principle - Transparency: to be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence. [↑](#footnote-ref-52)
53. CTS R7 requirement - Data integrity and authenticity: the repository guarantees the integrity and authenticity of the data. [↑](#footnote-ref-53)
54. CTS R9 requirement - Documented storage procedures: the repository applies documented processes and procedures in managing archival storage of the data. [↑](#footnote-ref-54)
55. CTS R10 requirement - Preservation plan: the repository assumes responsibility for long-term preservation and manages this function in a planned and documented way. [↑](#footnote-ref-55)
56. CTS R2 requirement - Licence: the repository maintains all applicable licenses covering data access and use and monitors compliance. [↑](#footnote-ref-56)
57. CTS R4 requirement - Confidentiality/Ethics: the repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms. [↑](#footnote-ref-57)
58. CTS R7 requirement - Data integrity and authenticity: the repository guarantees the integrity and authenticity of the data. [↑](#footnote-ref-58)
59. <https://www.rd-alliance.org/plenaries/rda-17th-plenary-meeting-edinburgh-virtual/repository-features-across-initiatives-0> [↑](#footnote-ref-59)
60. <https://rd-alliance.org/user/login?destination=group/node/44787/subscribe/og_user_node> [↑](#footnote-ref-60)