Open Data

Repositories



Depositing your data in a publicly accessible repository which assigns a globally persistent identifier ensures that your dataset continues to be available to both humans and machines in a usable format in the future. Funders and journals often maintain a list of endorsed repositories for your use. Still, choosing the best repository from such lists can often be daunting. Here, we offer some preliminary guidance on how to select a data repository.

Does your data contain **Controlled access repositories** personal or sensitive information that cannot There may be cases where openly sharing data is not feasible due be fully anonymized? to ethical or confidentiality considerations. Depending on what the Institutional Review Board approving your study said about data NO YES sharing, and what your participants consented to, it may still be possible to make your data accessible to authenticated users via a controlled-access repository. Is there a discipline Discipline-specific repositories specific repository for your dataset? Research data differs greatly across disciplines. Discipline-specific repositories offer specialist domain knowledge and curation YES expertise for particular data types. Using a discipline-specific repository can also make your data more visible to others in your research community. **Does your Institutional repositories** institutional repository accept data? Many institutions offer support providing repository infrastructure to their researchers for managing and depositing data. Institutional repositories that accept datasets provide stewardship, helping to ensure that your dataset is preserved and accessible. **Generalist data repositories** General data repositories accept datasets regardless of discipline or institution. These repositories support a wide variety of file types

does not exist.

and are particularly useful where a discipline-specific repository



Metadata



To aid discoverability, data should also be described using appropriate metadata; additional contextual information that allows others to understand your dataset. The content and format of metadata is often guided by a specific discipline and/or repository through the use of a metadata standard. Regardless of the repository you choose, when depositing your data it is important that you provide as much information as possible for your metadata record.

Versioning



Some repositories accommodate changes to deposited datasets through versioning. Selecting a repository that features versioning gives you the flexibility to add new data, restructure, and make improvements to your dataset. Each version of your dataset is uniquely identifiable and maintained — meaning others can find, access, reuse, and cite whichever version of the dataset they require.

Software



Software and code are important research outputs. In addition to using a version control system such as GitHub, you should deposit your source code in a data repository where it will be assigned a unique identifier. Using such a repository will ensure your code is openly and permanently available.

Data and code



Where you have both data and code, you should consider using a reproducibility platform like Code Ocean. Depositing your data and code in such a platform means that others can easily re-run your analyses, thereby promoting computational reproducibility.

Toolbox



- Re3Data
- (C) FAIRsharing
- (h) FAIR Repository Finder
- Making Your Code Citable



Caution!

Hosting your data solely on a laboratory website or as part of a publication's supplementary material hinders findability and reuse.



Caution!

Where you deposit your data will depend on any applicable legal and ethical factors, who funded the work, and where you are hoping to publish.

