Open Data
Repositories

Depositing your data in a publicly accessible, recognized repository which assigns a globally persistent identifier ensures that your dataset continues to be available to both humans and machines in a usable form 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 personal or sensitive information that cannot be fully anonymized?

NO — YES

1 Controlled access repositories

There may be cases where openly sharing data is not feasible due to ethical or confidentiality considerations. Depending on what the ethical board approving your study said about data sharing, and the level of permission granted from participants, it may still be possible to make your data accessible to authenticated users via a controlled-access repository.

Is there a discipline specific repository for your dataset?

NO — YES

2 Discipline-specific repositories

Research data differs greatly across disciplines. Discipline-specific repositories offer specialist domain knowledge and curation 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 repository accept data?

NO — YES

3 Institutional repositories

Many institutions offer support to their employees for managing and depositing data. Institutional repositories that accept datasets provide stewardship, helping to ensure that your dataset is preserved and accessible.

4 General data repositories

General data repositories accept datasets regardless of discipline or in- stitution. These repositories support a wide variety of file types and are particularly useful where a discipline-specific repository does not exist.

Metadata

To aid discoverability, data should also be described using appropriate metadata. 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 fill in as many fields as possible as this information usually contributes to the metadata record(s). In some cases, specifically where using a discipline-specific repository, the submission of metadata files alongside the data may be required.

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
- FAIRsharing
- 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.