

# Your go-to guide to making your data Findable, Accessible, Interoperable, and Reusable (FAIR)

So that you and others can get the most out of your data, it is important that you adhere to the [FAIR principles](#) to ensure your data are **Findable, Accessible, Interoperable, and Reusable** – whilst making your data openly available where it is safe to do so. This is no small task, so here are some ideas to help you get started:

## 1 Start with a management plan

An output management plan (OMP) is a useful starting point for collecting or creating data, software, research materials, and intellectual property. Creating an OMP before you begin your research, and updating it throughout the research cycle, will help ensure that your outputs are as open and **FAIR** as possible when your project is complete.

Some funders require grant-holders to produce a plan as part of their application for funding, and/or after funding has been secured.

You should consider:

- What outputs you will be creating or collecting, and how these will be documented
- What ethical or legal requirements, if any, apply to the outputs
- How you will organise, store, secure, and share the outputs
- What resources are required and who is responsible

## 2 Describe and document your data for humans and machines

Describing how your data were created, how they are structured, and what they mean is crucial to making your data **reusable**. As a general rule, someone who is not familiar with your data should be able to understand what it is about using only the metadata and documentation provided.

Good metadata is clearly associated with the dataset it describes and is available in a machine-readable format, such as text or RDF. Depending on your field of study, there may already be standards in place that will help guide how your data and metadata should be structured, formatted, and annotated.

## 3 Preserve your data

Data preservation helps ensure that your data will be **accessible** and **reusable** in the future.

Best practices for data preservation include:

- Backing up data files regularly
- Storing master copies of data files in open formats
- Validating preserved data files regularly
- Using more than one form of storage for data files
- Appropriately securing data physically, and/or on any network or computer on which they are held

### Bonus!

These practices also support reproducibility.



### Delve into the details

This is the first in a series of guides to help you be **FAIR**.

Advance your workflow with information on formatting your data in spreadsheets, selecting an appropriate repository, and openly licensing your data.

- [Spreadsheets](#)
- [Repositories](#)
- [Licensing](#)

### Toolbox

- [Cambridge Data Management Guide](#)
- [Australia ANDS Guide](#)
- [University of Manchester Research Data Management](#)
- [JISC Data Management Toolkit](#)
- [Frictionless Data Field Guide](#)
- [UK Data Archive Managing and Sharing Data](#)
- [Open Data Institute](#)
- [Digital Curation Centre](#)
- [Data Carpentry](#)
- [DMP Online](#)
- [Go FAIR Initiative](#)

# Spreadsheets

Spreadsheets are commonly used for data entry, organisation, analysis, and visualisation. By following best practices when using spreadsheets, you help ensure your data is **interoperable** and **reusable** for both humans and machines in the future.



## DO

- Keep your raw data raw; calculations and analyses should be done in a copy of the file
- Put variables in columns and observations in rows
- Give each column a descriptive heading that does not include spaces, numbers, or special characters
- Differentiate between zero and null values
- Validate your data
- Keep a separate txt file with a title and a legend describing your dataset, and outlining any steps you take to tidy your data
- Use a version control system and back up your files
- Export each data file in an open non-proprietary format such as CSV or TAB, with a name that appropriately reflects the content of that file
- Check your data thoroughly. Your data should receive the same care as your publications



## DO NOT

- Put more than 1 piece of information in a cell
- Use colour coding, embedded charts, comments or tables – your spreadsheet is not a lab book
- Include special (i.e. non alphanumeric) characters within the spreadsheet, including commas
- Use merged or blank cells
- Create multiple worksheets within a spreadsheet

## Metadata

Each spreadsheet should be accompanied by a data dictionary. A data dictionary is a separate file where each variable is defined, including units and ranges, and often includes other useful information for interpreting the dataset. By helping others (and your future self!) better understand your data, a data dictionary supports **reuse** and reproducibility.



## Toolbox



- [Data Curator](#)
- [Open Refine](#)
- [Good Tables](#)
- [Data Carpentry](#)
- [OSF Guides: How to Make a Data Dictionary](#)
- [Data Organisation in Spreadsheets](#)

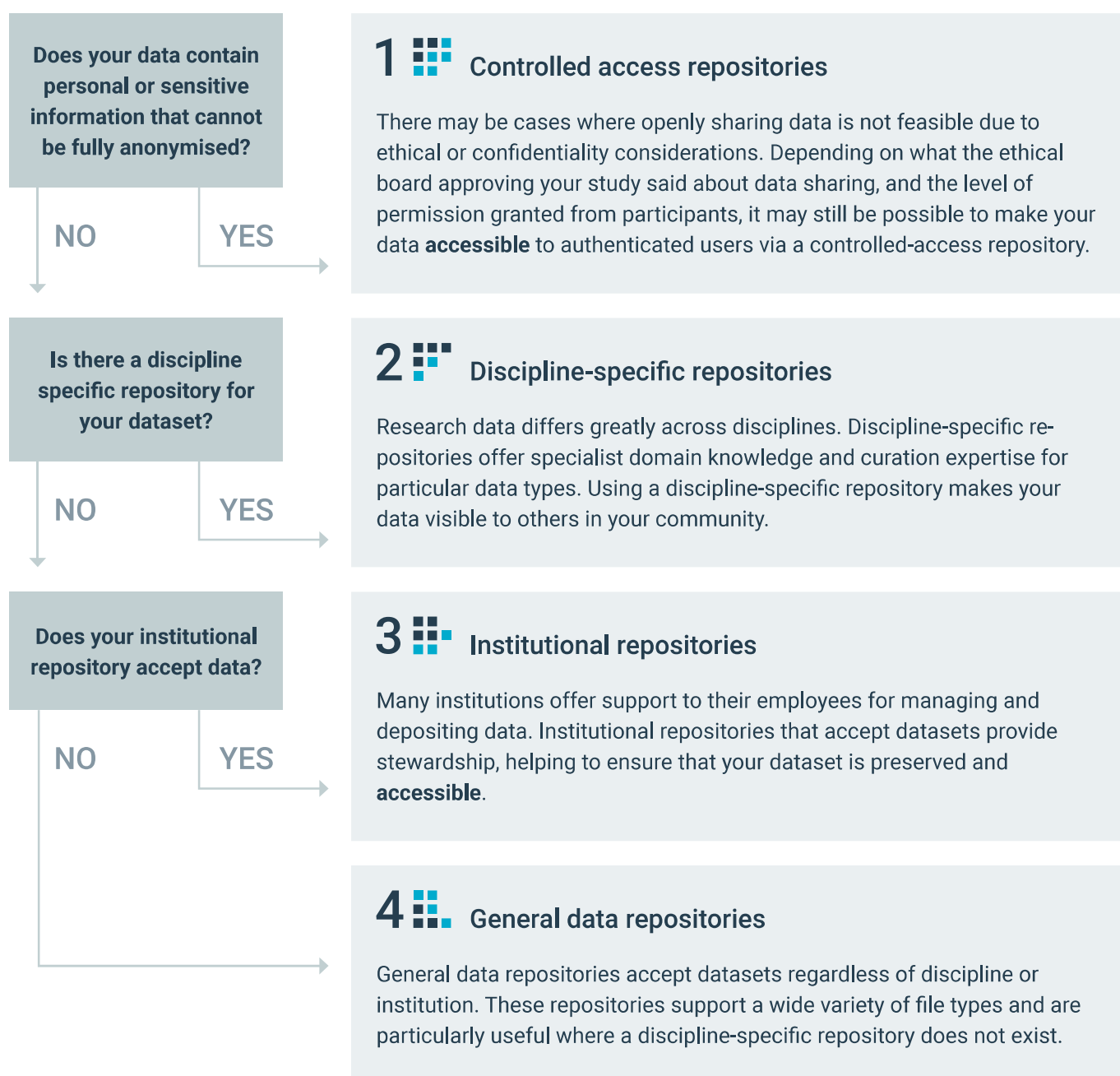
## Caution!

Different versions of spreadsheet software may handle data differently. Be especially cautious where your data contains dates or genes.



# Repositories

Depositing your data in a publicly accessible recognised repository which assigns a globally persistent identifier ensures that your dataset continues to be available to both humans and machines in a useable 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.



## 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** and **reuse** 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](#)
- [Research Data Support](#)
- [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 the journal you are targeting for publication.



# Licences

Data **accessibility** is defined by the presence of a user license. The license you select will determine the freedom with which others can **reuse** your data. When choosing a license, it is important that you adhere to any funder, repository, institutional, legal or ethical obligations.

## 1 Data



### CC0

**Creative Commons Zero** is ideal for openly sharing data – it has no restrictions on Reuse whatsoever. While CC0 contains no requirement for attribution, citing CC0 datasets is widely accepted and expected in science.

### Other CC licenses

If you find the use of a CC0 license is inappropriate for your data, you should consider the following CC licenses all of which require attribution in addition to further restrictions:

**CC-BY** – Prevents others from applying legal restrictions beyond the terms of the license to the licensed dataset.

**CC BY-SA** – Requires outputs derived from licensed dataset to also be licensed as CC BY-SA.

**CC BY-NC** – Prevents the licensed data from being used for commercial purpose.

**CC BY-ND** – Prevents the licensed data from being modified.

**CC BY-NC-ND** – Prevents the licensed data from being used for commercial purposes or modified.

**CC BY-NC-SA** – Prevents the licensed data from being used for commercial purposes, and requires outputs derived from licensed dataset to also be licensed as CC BY-SA.

### Caution!

NC, ND and SA licenses have implications for **reuse** and **interoperability**.

We suggest using a license that allows your data to be “as open as possible and as closed as necessary”.



## 2 Software

Making your software open source allows it to be freely used, modified, and shared by others. To ensure this is the case, you should consider using a license approved by the [Open Source Initiative](#). Popular OSI approved licenses include: [MIT](#), [GNU General Public License](#), and [Apache License 2.0](#).

### Dual licensing

It is possible to license your software under both an open source license (typically GNU GPL) and a proprietary license. The restrictions on the **reuse** of your software will then depend on which license the software is distributed under. Dual licensing allows you to potentially profit from your software whilst maintaining the benefits of open source licensing.

## 3 Applying a license

Once you've selected a license, you need to apply it. Most licenses include application instructions, so its best to follow these. Repositories also often support license application by allowing you to select a license from a pre-defined list on deposition.

### Caution!

Be aware of any licensing restrictions where your dataset contains data derived from a 3rd party.



### Caution!

Licenses cannot normally be revoked, and license conditions may differ with version.



## Toolbox



- [OSI Approved Licenses](#)
- [Choose an Open Source License](#)
- [CC License Chooser](#)
- [How to License Research Data](#)