Data Storage and Organisation

Contents

Data loss can be catastrophic for your research project and can happen often. You can prevent data loss by picking suitable storage solutions and backing your data up frequently.

Fig. 50 The Turing Way project illustration by Scriberia. Original version on Zenodo. http://doi.org/10.5281/zenodo.3695300

Where to Store Data

- Most institutions will provide a *network drive* that you can use to store data.
- *Portable storage media* such as memory sticks (USB sticks) are more risky and vulnerable to loss and damage.
- *Cloud storage* provides a convenient way to store, backup and retrieve data. You should check terms of use before using them for your research data.

Especially if you are handling personal or sensitive data, you need to ensure the cloud option is compliant with any data protection rules the data is bound by. To add an extra layer of security, you should encrypt devices and files where needed.

Your institution might provide local storage solutions and policies or guidelines restricting what you can use. Thus, we recommend you familiarise yourself with your local policies and recommendations.

When you are ready to release the data to the wider community, you can also search for the appropriate databases and repositories in [FAIRsharing](#), according to your data type, and type of access to the data. Learn more about this in the [Sharing and Archiving Data](#) subchapter.

Data Organisation

To organise your data, you should use a clear folder structure to ensure that you can find your files. We encourage you to use an existing template. An open source project created a quite complete one at [tonic-team/Tonic-Research-Project-Template](#)

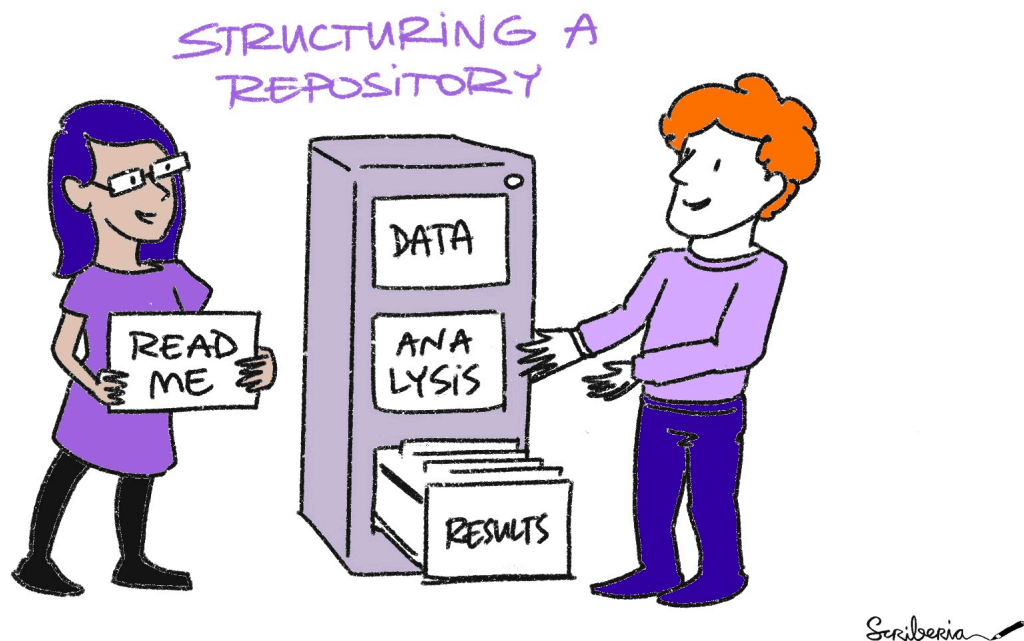*Fig. 51 The Turing Way* project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: 10.5281/zenodo.3332807.

- Make sure you have enough (sub)folders so that files can be stored in the right folder and are not scattered in folders where they do not belong, or stored in large quantities in a single folder.
- Use a clear folder structure. You can structure folders based on the person that has generated the data/folder, chronologically (month, year, sessions), per project (as done in the example below), or based on analysis method/equipment or data type.
- Avoid overlapping or vague folder names, and do not use personal data in folder/file names.

Data Organisation Examples

- Download this folder structure by Nikola Vukovic
- You can pull/download folder structures using GitHub: This template by Barbara Vreede, based on cookiecutter, follows recommended practices for scientific computing by Wilson et al. (2017).
- See this template by Chris Hartgerink for file organisation on the Open Science Framework.
- How to Organize Your Digital Files by Melanie Pinola.
- Project structure videos by Danielle Navarro (with slides).

More Information on Data Organisation

- [How to organise your data and code](#) by Rene Bekkers.

File Naming Conventions

Structure your file names and set up a template for this. For example, it may be advantageous to start naming your files with the date each file was generated (such as `YYYYMMDD`). This will sort your files chronologically and create a unique identifier for each file. The utility of this process is apparent when you generate multiple files on the same day that may need to be versioned to avoid overwriting. File names should be friendly to both machines and humans.

Some other tips for file naming include:

- Use the date or date range of the experiment: `YYYYMMDD`
- Use the file type
- Use the researcher's name/initials
- Use the version number of file (v001, v002) or language used in the document (ENG)
- Do not make file names too long (this can complicate file transfers)
- Avoid special characters `( )?\!@\*%{[<>` and spaces
- Hyphens `-` and underscores `_` can be used to separate related and unrelated chunks, respectively
- Keep in mind that some operating systems are case-sensitive, some are not
- Avoid personal data in file names

You can explain the file naming convention in a README.txt file so that it will also become apparent to others what the file names mean.

For further guidance on file naming:

- [Jenny Bryan's 'naming things' presentation](#) (or watch the [5 minute summary](#))
- [MIT's recommendations on File naming and folder hierarchy](#)
- [8 step guide on how to set up your file naming convention](#)
- [Project structure slides by Danielle Navarro](#)

File renaming tools

If you want to change your file names you have the option to use bulk renaming tools. Be careful with these tools, because changes made with bulk renaming tools may be too rigorous if not carefully checked!

Some bulk file renaming tools include:

- [Bulk Rename Utility](#), [WildRename](#), and [Ant Renamer](#) (for Windows)
- [Renamer](#) (for MacOS)
- [PSRenamer](#) (for MacOS, Windows, Unix, Linux)

Backups

To avoid losing your data, you should follow good backup practices.

- You should have 2 or 3 copies of your files, stored on
- at least 2 different storage media,
- in different locations.

Backups are ideally done automatically and should take into consideration your institute's guidelines. The more important the data and the more often the datasets change, the more frequently you should back them up. If your files take up a large amount of space and backing up all of them proves to be challenging or expensive, you may want to create a set of criteria for when you back up the data. This can be part of your [Data Management Plan](#).

Watch this video on [Safe data storage and backup](#) from the [TU Delft Open Science MOOC](#).