Wolfson Academic Skills
Research Data Management

This is the organisation, storage, sharing and archiving of all data produced during the research process. Many funders require that you manage your data in a specific way. In addition, it increases efficiency for you and ensures the preservation of data beyond the end of your project for others to reuse and mine for new findings.

What are data?

All research produces data regardless of academic discipline through collection, observation or creation. Analysis of data produces and validates your original research results. Data can take the form of archival material, audio files, code, documents, experiment results, field notes, images, interview transcripts, lab books, materials, physical samples, protocols, social media outputs, statistics surveys, videos and more.

Personal data is defined by the 1998 Data Protection Act as anything which can ‘reveal the identity of an actual living person’. You therefore need to consider whether the data you collect falls into this category and how you will deal with it. The easiest thing is not to collect it in the first place. But if you do, ensure you get informed and written consent, anonymise it, remove identifiers, aggregate results where there might be low frequency counts, remove outliers and use managed access repositories to store the data. You should report data breaches within 72 hours of it taking place.

Data Management Plan (DMP)

Many funders require researchers to manage data. To ensure you are following best practice you should create a DMP before starting your project.

The plan should:
- Outline data to be collected
- Detail how it will be managed
- Cover the lifetime of the project and beyond
- Include specifics on data type, storage, backup, metadata, sensitive information and sharing data.

Think about best practices for dealing with data at every stage of the lifecycle to develop a robust strategy for data management.

There is a whole section on writing a DMP on the Cambridge’s Data Management site but this image summarises the stages involved.

The Research Cycle adapted by University of Denver from the Research 360 Project.

Lee Library Wolfson College

Turn over for more tips
Organising your data

Once you create data and files, they can quickly become disorganised. You need to establish a system that allows you to access them quickly and avoid duplication. Start by developing a logical folder structure.

It is important to name your files so that they are:

- consistent
- meaningful to you and others (if you work on a shared drive)
- findable

For physical samples, you could create maps of your storage system, reference samples in notebooks, and add notes to the samples themselves.

With digital files, think about using the following pattern: prefix (for document type e.g. report, notes, essay)_document title_version_dateyyyymmdd

Keep folders structured similarly, using meaningful titles plus dates where practicable to divide up work. Nest folders to keep each level to of storage to a minimum. Having 50 folders on your desktop is just as confusing as having 50 documents.

Metadata

Where possible include a further file containing metadata (data about data) in the same folder as your data. This will allow you to add context to your data so that you and others can understand it in the short, medium, and long-term.

Best practice is to:

- Write it as a plain text file
- Call it: README
- Include general information: title, authors, date of collection
- Provide an overview - short description of the data each file contains and date it was created
- Be explicit about how the data can be shared; is there a licence or are restrictions placed on the data?
- Describe methodological information: how was data collected, generated and processed?
- Include data-specific information: variable list (including definitions) for tabular data, units of measurement, definitions for codes or symbols used to record missing data.

Backing up your data

Consider what would happen if your data disappeared because you lost a USB stick, spilled coffee into your laptop, had it stolen, files were accidentally deleted, or there was an incident at your department. This happens all the time to individuals and large organisations. Options for back up include:

- Digitising any analogue data
- Departmental back-up
- External drives
- Online back-up
- Automated services

Your strategy should be guided by considering: what you are willing to lose, what is crucial to your research and how often does it change? The more it changes, the more often you need to back it up. Undertake to back it up on a weekly basis (as a minimum) if you are working on it regularly.

Be cautious with cloud solutions and check the small print to ensure that you are not giving the host permission to reproduce, modify, and publish content. Check the UIS pages on cloud solutions https://help.uis.cam.ac.uk/service/storage/storage-overview for more information.

ALWAYS have data backed up in two different locations. There is no point saving data to a USB if you store it with your laptop; they are equally vulnerable.

You should aim to store it for at least ten years in a suitable repository and link it to your publication(s). Data can be uploaded directly through Symplectic www.data.cam.ac.uk/upload

When preserving data ensure that it is:

Findable - includes metadata and a unique identifier
Accessible - protocols are open, free, and universally implementable
Interoperable - uses open source file formats
Re-usable - there is clarity on usage licences

More information is available from the Research Data Management Team:

https://www.data.cam.ac.uk/