Data collection tips and resources

We know managing data collection can seem daunting for many early career researchers, so we have spoken to our data champions and put together a list of data collection tips and resources that we hope will help.

Collection

- Have a clear protocol in place, from the beginning of your project, outlining how you will save both your data and metadata, to ensure you will be able to quickly and easily access any portion of your work that you might need at a later point for further analysis, consultation, etc.
- Use the clouds offered by Universities or institutions and hard-drives to have backup your work.
- Consider using an **electronic laboratory notebook**.
- Encrypt or use codes when saving sensitive data and keep a log of each step.
- Check ethical requirements if acquiring data from patients as soon as possible.

Ensuring reproducibility

- Make sure you continually update your outputs management plan, taking into account any changes in the direction of your research and the outputs you anticipate generating.
- Keep a log outlining how your data were created, how they are structured, and what they mean.
- Make sure you use **metadata and data standards** appropriate for your field.
- When creating new versions of your data, record what changes have been made and give the new files a unique name.

Collaboration

- Make use of available software to help to facilitate collaboration and data sharing. These include **Github/Gitlab**, **Labkey** and **Jupiterlab**.
- Make sure all your collaborators agree, from the outset, how the data and other outputs from the project, will be shared and how credit will be assigned.
- Setup a clear pathway for communication and feedback, e.g. a slack channel or other online platform.

Data re-use

- Check to see if complementary data, helpful to your project, exists using search engines such as **google dataset search**, **re3data.org** and **Datacite**, which harvest information from publicly available digital repositories.
- Data providers are another valuable source of UK and international data. Examples include: **UK data service**, **Office for National Statistics**, **World bank open data** and **UN data**.
- Check the license associated with the data to verify what you can and can't do with the data.

Further help

- Check the dedicated Research Data Management support website of your university library.
- Contact your institutional Research Data Support for questions.
- Join a community or an initiative such as an **Open Science Community**, **ReproducibiliTea**, or **The UK Reproducibility Network (UKRN)**.

Toolbox

- UKDS Guide on Research data management:
  https://www.ukdataservice.ac.uk/manage-data.aspx
- The Turing Way – open source community-driven guide to reproducible, ethical, inclusive and collaborative data science:
  https://the-turing-way.netlify.app/
- UK Reproducibility Network resources:
  https://www.ukrn.org/resources/
- Data and metadata standards:
  https://fairsharing.org/standards/
- REDCap – a secure web application for building and managing online surveys and databases:
  https://www.project-redcap.org/
- Git and GitHub for beginners – crash course:
  https://www.youtube.com/watch?v=RGOj5yH7evk
- Software and Data Carpentry lessons:
  https://software-carpentry.org/lessons/
  https://datacarpentry.org/lessons/
- Code Refinery – Training and e-Infrastructure for Research Software Development:
  https://coderefinery.org/