

MOORE ABOUT ... DATA MANAGEMENT



BETTY AND GORDON MOORE LIBRARY

RESEARCH DATA MANAGEMENT RECAP

What is data?



Research data is any information used or created during the research process. Although the term 'data' is common in the scientific disciplines, those working in the arts and humanities do not always think of the information they work with in this way. It is important to realise that the term covers a wide range of material from statistics and lab results to historical documents and interview transcripts.

The data involved in each project will be different depending on the nature of the research being undertaken. The approach taken to data management will also vary depending on the information in question and any guidelines provided by funders.

Why should it be managed?

Good management of data is an essential part of any project – it can help to make the work run smoothly and saves time after the project has finished. Organising data makes it easy to find the information that is needed at any stage. Many funders will now demand to see a plan of how data will be managed before releasing any money for a project.



As the use of Open Research practices becomes more widespread there are other reasons for managing data. The increased emphasis on sharing the data which underpins any output means that information needs to be prepared for sharing with those outside the immediate project. Not only does this information help others to build upon the original work but it means that the research can be verified if necessary – something of increasing importance in an age of fake news and misinformation!

RESEARCH DATA MANAGEMENT RECAP

How is data shared?



With the rise of Open Research more and more funders, researchers and their institutions are realising the benefits of sharing data. Different research funders have different requirements but this is usually done by depositing the data in a repository. There are repositories dedicated to different disciplines, specific types of data and those which will accept any type of information so there is something to suit every need.

When sharing data it is important to think about how it can be reused by others. The F.A.I.R. Principles for the management of data can be used as guidelines to help with this. Data should be:

- **Findable** – make sure that the data is discoverable using accurate metadata and a DOI.
- **Accessible** – ensure that as many people as possible can use the data by using open source systems and universally recognised formats.
- **Interoperable** – use standards and formats for data which work across different systems.
- **Reusable** – make sure that data can be used and built upon by others by attaching an appropriate licence such as Creative Commons.

Repositories should follow these guidelines as far as possible, especially when it comes to keeping the data accessible. A good repository will commit to long term preservation of and access to the material it holds, providing a persistent link so that researchers do not have to keep updating links in their online profiles.

It is worth remembering that some repositories charge for the storage of data, particularly if the data set is very large. This cost should be factored into any grant application or other research funding to avoid a nasty surprise at the end of a project.

RDM PRINCIPLES

When it comes to good data management there are many areas to think about. Most research funders provide specific guidance on what they expect but in general you should think about the points below.



ORGANISATION

Good data management starts with the organisation of information. Researchers need to implement a system for both their digital and physical information which allows them to find what they need quickly and easily. This is especially important in large groups where different people may need to access a range of data. For digital information a file naming strategy should be considered. This helps to keep names consistent across teams and projects.



STORAGE

Data will need to be stored both during and after a project so that it can be accessed as needed. It is particularly important that there is a plan to back up any data that is created to avoid losses. Researchers should think about the method of backup used, how often they will update this and the amount of information that will need to be backed up at any one time. They will also need to consider a storage location and whether this should be physical or Cloud based.



PERSONAL DATA

Many projects deal with personal data in a variety of forms. This information needs to be carefully managed and protected. When gathering this information researchers should ensure that they are explicit about exactly what will happen with the information and how it will be stored, used and shared.



SHARING

Research funders now mandate that data from a project is shared more widely. This should be done via a suitable repository using adequate description and an appropriate licence so that others know how the information can be reused.



This work is licensed under a Creative Commons CC-BY 4.0 license by Claire Sewell. Betty and Gordon Moore Library, Cambridge University Libraries.