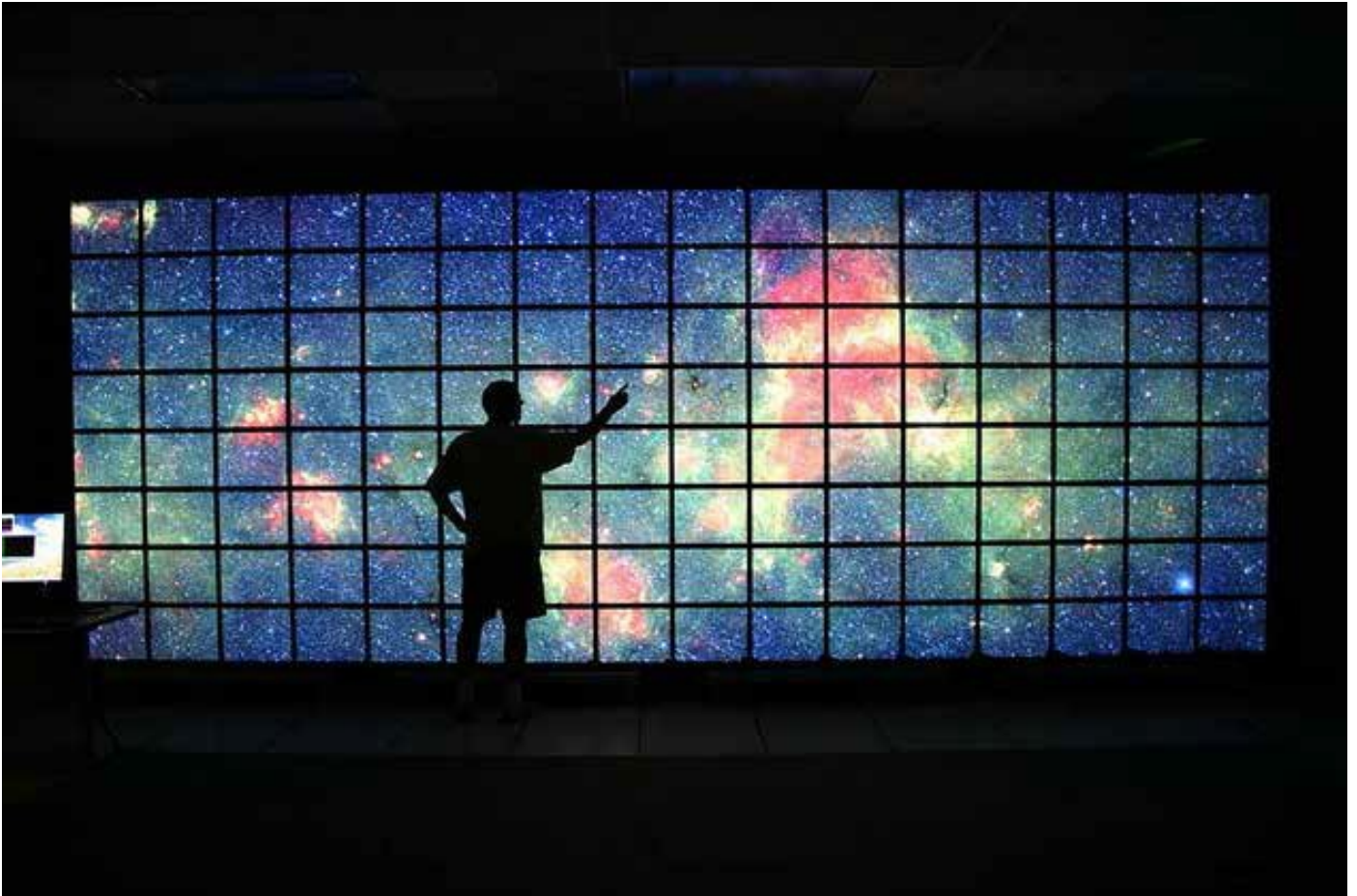


Bitesize Guide: Research Data Management

Research data (n.) *“Recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings. Although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.”*¹



¹ <https://www.epsrc.ac.uk/about/standards/researchdata/scope/> Other definitions of research data exist and the nature of research data can vary depending on the discipline. For example, AHRC data management requirements apply primarily to “significant electronic resources or datasets”, rather than the primary sources which a researcher may consult in an archive or library.

Why manage your research data?

Research data management is an important part of good research practice.² Alongside considerations of research ethics and intellectual property, researchers have a responsibility to plan, organise, store and, where appropriate, share their data with other researchers. Managing research data helps you to maintain transparency and accountability for your research, enhances data security and reduces risk of data loss. Good research data management ensures the long-term value of research and as such benefits the researcher, the University, the academy and the wider community.

Many research funders recognise the importance of good research data management. Research Councils UK's Common Principles on Data Policy³ states that 'Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.' Research funders often have mandates to ensure compliance with their research data policies, and failure to comply may lead to sanctions, such as withholding payments from current awards and an inability to bid for future awards.

There are a number of stages involved in managing your research data, including planning how you are going to manage the data; organising and describing your data - so that you and others can use and find it; robust and secure storage and backup - to ensure confidential data is safe and nothing is lost; and sharing and preserving long-term access to your data.

Planning research data management

All researchers, including staff and postgraduate research students, should write a Data Management Plan (DMP) prior to the start of a project to ensure that all issues relating to the management of your data are fully considered at the outset. Research funders may require a DMP to be submitted as part of a grant application. The length of a DMP should be appropriate to the nature and scope of the research you are undertaking, but it will typically cover the following:

- Data collection: what kind of data will you collect or create?
- Description: what documentation and metadata will describe the data?

- Ethics and Legal: how will you manage ethical and intellectual property issues?
- Storage: what data will you store and where? How will you manage backups?
- Sharing: how will you share the data and are there any restrictions on this?
- Preservation: How will you ensure the data is preserved for the longer term?

To address some of these areas, you may need to consider including additional costs in external funding bids; for example if you need additional staff time or specialist software to prepare data for sharing. You should speak to the Pre-Award Team in Research & Innovation Services about this to ensure that these costs, where eligible, are included in the bid.⁴

A DMP will also assign roles and responsibilities for management of data and you should ensure that you review and, where necessary, update the document throughout the project. For participatory research, where data is generated collaboratively, ownership and therefore management of data may need to be negotiated alongside other aspects of the collaboration. This should be reflected in the DMP.

Further help and guidance: The Digital Curation Centre hosts the DMPOnline tool (<https://dmponline.dcc.ac.uk/>) which provides templates to help you create a DMP. Both generic and funder-specific DMP templates are provided.⁵

Organising and describing your data

This is an often overlooked but crucial part of data management. Good organisation ensures you can easily find data when you need to, and describing your data means it can be more easily understood and reused by others. This includes consistent folder and filename conventions applied across the researchers involved in the project:

- Aim for filenames which are concise but meaningful
- Files are often sorted alphabetically, so you may want to put a date, author, or subject at the start of the filename so you can
- Adopt a consistent approach to version control (e.g. v0.1, v1.1 etc.)

² Research is defined in accordance with the Frascati Manual and the REF 2014 guidance as "a process of investigation leading to new insights, effectively shared". This definition includes work of relevance to commerce and industry, but excludes routine testing and analysis: <http://www.ref.ac.uk/2014/pubs/2011-02/>

³ <http://www.rcuk.ac.uk/research/datapolicy/>

⁴ <https://one.northumbria.ac.uk/service/ris/Pages/Research-Resources/Applying-for-Funding/Developing-Costing-Proposals/Research-Bid-Timeline-Countdown-to-Submission.aspx>

⁵ To use the DMPOnline tool, you may be asked to register for an ORCID. This is an international, community-driven effort to create and maintain a register of unique researcher identifiers and the University encourages all staff to register. See: <http://library.northumbria.ac.uk/info-researchers/orcid>

Describing and documenting your data is important so that others can find, understand and use it. Data documentation “explains how data were created or digitised, what data mean, what their content and structure are, and any manipulations that may have taken place”. Metadata are a subset of data documentation. Metadata “provide standardised structured information explaining the purpose, origin, time references, geographic location, creator, access conditions and terms of use of a data collection”⁶. You can, for example, create and review metadata in Microsoft Office by viewing and editing the Properties fields: Title, Tags, Comments, Status, Company, Author etc. Other software may have different metadata creation functionality.

Further help and guidance: The University’s Records Management Office has produced further guidance on versioning, organising drives, and filenames: <https://www.northumbria.ac.uk/about-us/leadership-governance/vice-chancellors-office/legal-services-team/records-management/records-management-guidance/>

Data selection, storage and backup

It is important to appraise your data and select what to retain and what you can or should discard. Not all digital data can be kept indefinitely: storage costs increase as data size increases, and it may not be appropriate to retain some types of data for legal or ethical reasons. You can use the checklist of criteria below to help you decide, but you should fully consider this when writing your DMP:

- What data does my funder and/or the University require me to keep? Consult the relevant policies to ensure you are compliant.
- Is there any legislation which requires me to retain or destroy certain data? The UK Data Protection Act and the EU’s General Data Protection Legislation are relevant here (see below).

Research with personal data: On 25th May 2018, the General Data Protection Regulation comes into force in the UK. <https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/>.

The GDPR brings into play additional rights, restrictions and responsibilities around using personal data. The main difference for research and researchers is the requirement for a greater degree of *transparency* around how data is being processed, as well as additional *safeguards* to ensure personal data is not retained for too long, and is stored securely.

To find out more and understand how you can be compliant with GDPR, read the new sections in the [Research Ethics & Governance Handbook](#).

- Is your dataset unique/particularly valuable for scientific, historical or cultural reasons? For example, if the data relates to a one-off event that would be difficult to repeat.
- Are there any issues involved in sharing the data? For example, intellectual property rights and licensing issues.⁷
- Is the cost of keeping the data acceptable and is funding in place? Storing and sharing data costs money and this can be a factor in deciding what to

Further help and guidance: The Digital Curation Centre hosts a five-step guide on to help you decide what data to keep: <http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep>

discard.

It is vital that you ensure your data is stored securely during your research project. While storage on a USB stick or external hard drive might be convenient in terms of portability, it leaves data vulnerable to loss or interference. Sensitive or confidential data should be encrypted and the University recommends that all research data is stored on a networked drive to ensure regular backups are made. Storage on unsupported cloud-based services (e.g. Dropbox) is not recommended due to potential security risks, but the University is in the process of procuring cloud-based storage which will provide a safe and secure file storage and sharing solution. Further guidance will be issued once this is available.

Preserving and sharing your data

Funders frequently mandate that data should be preserved and accessible for significant periods of time: for example, EPSRC mandates that research data should be accessible for at least 10 years. The University’s RDM policy is based on the research records retention schedule⁸ which stipulates varying periods of retention dependent on the risk inherent in the project and/or funder requirements, whichever is the longer. This may mean preserving data and making it accessible for three or more years after project completion. For this reason, providing access to data via a project website is often not

Further help and guidance: There are a number of subject and funder data repositories available: <http://www.re3data.org/search?query=> search service for a range of subject data repositories; figshare.com/ for all disciplines; ukdataservice.ac.uk/ for social sciences (ESRC funded); data-search.nerc.ac.uk/ for environmental sciences (provides access to all NERC data centres); datadryad.org/ for life sciences.

⁶ <http://www.data-archive.ac.uk/create-manage/document/overview>

⁷ Normal procedures around management of intellectual property apply: <https://www.northumbria.ac.uk/media/6742668/intellectual-property-policy-june-2015.pdf>

⁸ <https://www.northumbria.ac.uk/-/media/corporate-website/documents/pdfs/about-us-corporate/legal-services-team/3,-d,-research.pdf>

adequate as these are frequently abandoned or poorly maintained after the project is finished. To ensure your research data is preserved and remains accessible over the longer term, you will need to consider depositing it in a dedicated data repository.

The University is currently developing its infrastructure for data deposit and sharing and this guidance will be updated once this is in place. Staff are advised to consider external subject data repositories (see list above). Please contact us (researchdata@northumbria.ac.uk) for further advice and guidance.

Further Help and Support

Northumbria's RDM policy and webpages:
northumbria.ac.uk/research/research-data-management/

Each Faculty has appointed an academic lead who can be contacted to discuss general research data management issues:

ADSS: Sarah Ralph

BL: Ian Charity

EE: Gobinda Chowdhury

HLS: Leigh Riby

For further information about research data at Northumbria please email: researchdata@northumbria.ac.uk

