Surrey Data Management Plan (DMP) Guide

Contents

1. What is a DMP and how do you evaluate it?

A DMP is a 1-2 page roadmap for **how you will collect/measure, analyse/process, document, protect, share, store and archive research data from your project**. It's best to write a DMP before a project begins and certainly before any data generation. It must be clear who is responsible for creating it and reviewing it regularly throughout the projects lifetime making updates as necessary.

There are five main areas a DMP should cover:

- **Files:** What type of data or files will be created or collected?
- **Documentation:** What documentation and metadata will be created?
- **Storage and security:** How will data be stored and protected?
- **Sharing:** Will all/part of the data be shared? If so, how, where, and when? If not, why?
- **Archiving:** How and where will data be archived/preserved for future re-use?

File types and data creation

This section should outline what types of files the project is creating.

Does the plan:

• List the characteristics of the data to be generated (e.g. quantitative/qualitative, numeric, text, audio, video, surveys, code/models, specimens, reuse of any secondary/existing data, etc.).
• How will file names/folders be controlled and what about file versioning?
• How is data consistency and quality to be controlled, e.g. calibration, repetition, peer review, etc.?
• Any methodologies or software to be used for data creation and subsequent analysis?
• What are the expected volumes (e.g. Bytes) or numbers of files, participants or surveys, etc.
• Include the file formats/software and if they are open or proprietary – this is relevant for the sharing and preservation sections.

Keep these files and formats in mind when creating the remaining sections for potentially unaddressed issues, like storage requirements, formats less suitable for long term preservation, and potential personal data and sensitivity issues.

Documentation and metadata

Is the proposed documentation enough to ensure someone else could easily understand, verify/replicate conclusions of the project, and potentially re-use/extend the data sometime in the future?

This could include:

- Outlining what explanatory documentation will be created for defining variables or units of measurement, e.g. data dictionaries, user guides, code books, documents, etc.
- Describing workflows for systematic capture of study information.
- How you will add to, review/update, and maintain the data and documentation.
- How you will handle non-digital documentation like lab notebooks, specimens, or written field notes.
- Are there any relevant disciplinary standards for documentation and metadata you could use?

Storage and Security

The two biggest risks to research data are *unauthorized access* and *accidental loss* from human error or technical failure. This section should outline how these risks will be mitigated during the project. In almost all cases you should be using University storage (e.g. OneDrive, Sharepoint) as your main storage space for your research data. University storage is managed by IT services and hence regularly/automatically backed up with the latest malware and anti-virus protection and secure access controlled with passwords.

Some projects may require data be collected in the field or using tools with other storage options (e.g. audio recorders with built in storage, online collection tools like Qualtrics or Gorilla, etc.) In each of these cases, the plan should outline how and when data will be moved onto secure University storage, including the timing for deletion from non-university storage. If you are using highly controlled (e.g. personal/special category data) or commercial data, you should outline the terms agreed by the data owner to keep your data safe and state who the data owner is.

Look for these things:

- Outline of where the data will be stored at every stage of collection, processing and analysis and who is responsible for it.
- Description of how and when data will be transferred, including deleting data off collection tools/storage.
- Any legal or ethical issues with the data, e.g. identifiable data, copyrighted materials, etc. How will you protect personal data? For example, transferring data, using secure university storage and use of key files kept separately, or transforming/de-identifying/anonymising the data after collection.
- Has an ethical review been conducted?
- Identifying who will have access to all or just part of the data, e.g. just you and your supervisor? A team? An external collaborator/company with a Collaboration Agreement? How will they ensure secure data transfer/access to the data? What's preventing data from being accidentally overwritten if accessed simultaneously by many?

- Any special storage (e.g. Trusted Research Environment (TRE)) or computing requirements (e.g. space, software, equipment or instrumentation) the project may have that hasn't been identified.
- Description of how you will securely store and maintain any non-digital data e.g. physical items/specimens.

Sharing data

Surrey is committed to transparency, reproducibility, and the value of quality research data. In line with funders, our Research Data Management procedure encourages researchers to share data as openly as possible. Additionally, you may be asked by journals to share the data underpinning your articles.

Supervisors and PGRs should have a conversation about what data may or may not be shared from their projects and outline the plan in this section. A popular benchmark for sharing data is the FAIR Principles - GO FAIR Is it Findable, Accessible, Interoperable, and Reusable?

This section of a DMP should outline and justify what parts of the data can be shared publicly, shared with restrictions, or not shared at all.

Things to include are:

- Consent forms that don't prohibit sharing/retention of de-identified/anonymised data and allow for archiving and future re-use of the data by others.
- Outline of what parts of the data can and cannot be shared (e.g. survey results can be publicly shared, but interview data will be deleted once transcribed or restricted to only researchers upon request after a review/registration process).
- Is there non-digital data that needs to be made available? How will that be handled?
- If the data cannot be shared, explanations of why (e.g. don't own, national security, copyrighted, no permission, etc.).
- Outline of how data will be transformed to make it more shareable/reusable (e.g. de-identify or aggregate it or convert it to an open format).
- A plan for how/where/when data will be shared/made accessible, such as depositing it in a repository.
- You may refer to a time frame for sharing, such as alongside publication or at the project close. You may refer to your funder's required timeframe for sharing data.
- You should state the terms and conditions for access such as an intended copyright license for the shared data About CC Licenses - Creative Commons (e.g. CC-0, CC BY, CC BY-NC-SA).

Archiving

Regardless of whether data is shared, it must be retained for a minimum of 10 years. Here we would expect you to include what data will be retained, where it will be stored, and who will oversee its safe keeping. Additionally, you can identify actions to increase the long-term viability and reusability of the data both from a technological point of view (how likely is someone still able to open this file in 10 years using what software/hardware?) and from a documentation point of view (will they be able to correctly understand and reuse this data?).

We would expect to see the inclusion of:

• What data should be archived. This should be anything that underpins the conclusions/main findings of the project and any published works.
• What explanatory and contextual documentation will be included to facilitate data verification and possible future reuse.
• Whether you will transform your data into an open format for preservation, keep it in its original format, archive raw and/or processed data.
• If you are going to share your data in a repository, you may reference the repository's preservation policy.
• Will a persistent identifier (e.g. Digital Object Identifier (DOI)) be assigned to the data?
• Are any additional costs anticipated, e.g. for storage and data curation or cataloguing?

2. Resources

DMPonline is a free online tool to help construct a DMP. It's geared towards those applying for grants, but there is a generic DCC template and many funder templates that you can use for any project. They also provide some example DMPs written for specific funders.

Open Research has guidance on their Library website pages Data management plans | University of Surrey about DMPs and regularly run DMP workshops via the Doctoral College and Staff Development. Additional information is provided on the Library Research Hub - Home.

If you or your supervisor have any questions, please contact Open Research at openresearch@surrey.ac.uk or come along to the *RDM drop-in session on Tuesday mornings (8:30-13:00) on Level 1 of the Library opposite the Entrance Gates. They also offer a DMP review service.*

All the relevant University policies that are being followed on the project should be listed in the DMP Policies and procedures | University of Surrey.

3. Frequently Asked Questions

What are the advantages of doing a data management plan?

Principles: Surrey is committed to fostering an open research culture Open Research | University of Surrey. DMPs are a useful tool to help increase openness, transparency, and reproducibility of research data. See Open data | University of Surrey.

Procedure: The University's Research Data Management procedure research-data-management-procedure.pdf expects DMPs for all research projects, including PGR projects.

Professionalisation: Most funders in the UK, Europe, and North America require DMPs as part of their applications. This is good practice if a PGR is continuing in academia and provides an opportunity for PGRs and supervisors to discuss open research practices within their disciplines.

Ethics, GDPR, and Intellectual Property: A DMP provides another chance for supervisors and students to discuss how to appropriately manage any personal and sensitive/special category data. See Ethics | MySurrey student website | University of Surrey.

What do you mean by "data"?

Great question. *Anything that underpins your conclusions or could be re-used by someone else in your field*. Imagine if someone wanted to replicate the project – what files, scripts, code, and documentation would they need to achieve that? That's the data! You can find a more formal definition in Section 3 of our Research Data Management Procedure.

I'm not sure I have any "data". I work in an arts/humanities/theoretical area.

For these areas DMPs look more like project management plans for file management, storage, sharing and preservation. Plans in these areas likely will focus on organizing digital files (sources, research notes, and drafts of publications) and any physical materials you'll be using. This could include using software (e.g. reference managers) to link notes and sources or even something as simple as a file naming convention or version control to keep track of types and the updating of files.

DMPs should outline steps to reduce the likelihood of accidentally losing digital work, like using University cloud storage on OneDrive or creating multiple copies or doing regular back-ups. For the sharing part of a DMP, likely the "data" may be secondary/existing so will already be shared in publications, copyrighted, or accessible from other sources (e.g. archives, literature or museums) so comprehensive referencing is important. Finally, you will want to take some steps to keep your project files (which is also data) safe as part of a broader retention of your corpus of work.