

Doctoral Data Management Plan

guidance for supervisors

All doctoral students are required to write a Data Management Plan (DMP). This should be alongside their Academic Needs Analysis for students who do ANAs and at the latest by their first progression review. The DMP is a living document and should therefore be reviewed regularly throughout the research project, at a minimum at each progression review.

Training and support for Data Management Plan development

All new doctoral students are required to take the Data Management Plans for Doctoral Students course on Blackboard. Students will be issued with a certificate when they complete the module.

The Library Research Data Service also provides guidance through their webpages, the DMP template and guidance for students can be found in the course and on the library's webpages at: <https://library.soton.ac.uk/thesis/data-plan>

The DMP template has a coversheet and eight questions that the students have to answer. It is not expected that students will be able to answer all questions comprehensively at the start of their research. A DMP is a living document and some answers will change as the research progresses and the student becomes more familiar with their data and disciplinary norms.

Coversheet:

About the Research

Most of this section is purely factual. Where students need ethical clearance, the DMP should be written with the ethics application in mind to ensure that the ethics and DMP do not contradict one another.

About this plan

The student should indicate how regularly the plan will be reviewed. Review times will differ with the nature of the research and what stage the research is at, for example a plan that had been written before data collection has begun should ideally be reviewed shortly after collection commences. Minimum requirement is at every progression review.

The student should identify any further training or support that they might need to implement the DMP. If any answers imply that resources (e.g. data storage) will be needed beyond normal University provision, these resources should be clearly specified in the context of the project.

Version Table

As this is a living document, a version table is included so students and supervisors can more easily keep track of which DMP version they are reading.

1. Project Description:

The student should write a short paragraph summarising their project's research questions and data needs. They should give a context for their research, for example, is it part of a larger programme and if so, consideration should be made about how the Doctoral DMP aligns with a wider programme DMP.

2. What policies will apply to your research?

This section should include links to local, national or international policies and legislation that must be complied with during the course of the project with regards to the management of the research data. The RDM website and the guidance notes list some of these which might apply, and more are provided in the online course.

The student should also summarise any contractual requirements (funder, industrial partner, collaborators) that relate to the management of their research data in this section.

3. What data/research material will you collect or create?

This section should include the following information:

- The type of data that will be used (qualitative / quantitative).
 - If the research is theoretical, students should discuss the material they are consulting.

- If creative material is to be explored or developed, students should consider what research materials will be used during the developmental steps.
- If software is to be developed, a good answer to this question should indicate:
 - The type of software e.g. script, model, module, library, prototype application, software product
 - programming language that will be used
 - whether the software will be uniquely fitted to the project or potentially useable in new contexts
- The formats of the data that will be collected.
- The volume of data that will be generated. This is to help them to plan storage needs with you if they are generating large volumes of data.

The student should provide a brief summary of the methods of data collection.

4. How will your data/research material be documented and described?

Students should show awareness of research diaries, lab notebooks and so on for documenting their research journey and recording the “what, where, when and how” of collecting data or research material.

The online course suggests schemas that might be applicable in their disciplines to use to collect the relevant descriptions.

The student should outline how they will organise their folders and files, the file-naming convention that they will use and how they will keep track of versions.

We are encouraging doctoral students to document their data in a readme file. We are also encouraging them to fully annotate any code that they write. For disciplines where code is being developed, they might specify that they will use GitHub or Mercurial for version control and where the upstream repository will be hosted.

5. How will you deal with any ethical and copyright issues?

In this section the student should address any issues regarding:

- Who else has a right to see or use this data?
- If the data is personal, sensitive or commercial how the student will safely share data within the supervisory team?
- How the student will anonymise data during research (if appropriate)?
- If the student is using secondary data, what restrictions are there on re-use and how will the student deal with these?
- How will the student destroy any personal, sensitive or commercial data which cannot be stored beyond the end of the project?
- If the data could be considered high value and/or vulnerable, for example, if the data likely to attract “hacktivists”, how could this be mitigated?

6. How will your data/research materials be stored, and backed up?

The student should demonstrate that they have considered:

- the safe storage of their data;
- storage options for high volume data;
- back-up of their data;
- security of their data.

We recommend to them that they store their data on the University’s managed storage as the primary location for their data, for more information see [researchdata storage](#) We also provide advice on how to use [safesend.soton.ac.uk](#) as a secure file transfer service

Valid reasons for alternative locations of data storage include:

- Use of non-digital data (consider digitising consent forms).
- High volumes of data (>5 TB).
- Use of highly sensitive data that requires use of a non-networked computer.
- Use of data being held by a third party.
- Working in the field.

If alternative data storage options are required a back-up schedule should be specified.



7. What are your plans for the long-term preservation of data/research materials supporting your research?

This section should cover the selection of data to be retained and deleted at the end of the project. It should state, at a minimum, that the following data will be preserved long-term:

- Data underpinning publications or results reported within their thesis;
- Data that would be impossible or prohibitively expensive to re-create;
- Data that are of value as a resource for future research;
- All relevant coding that would be necessary to reproduce the results from the raw data;
- If data are not being retained there should be justification for this.

Students should state the formats that the data need to be saved in for future preservation.

We recommend keeping raw data and data processing code rather than results such as graphs and tables. The student should also identify files that do not need to be retained at the end of the project (e.g. temporary files, intermediate files, audio/video recordings once transcripts are available).

Research data and material should not be retained post-project on networked My Documents, research filestore or OneDrive or solely on external hard-drives or laptops; none of these are suitable for long-term storage. Nor should data be stored on a project website or 'on request from the author'.

Data archives should be used for post-research storage. Ideally a disciplinary data archive should be used, for example Dryad, UK Data Service Archive, Genbank, EMBL, or a widely used general data archive such as Zenodo. If there are no discipline-specific data archives the University of Southampton Institutional Repository would be an appropriate archive. Using an archive has the additional benefit of allowing access to the material to be managed by the archive. Data archives can provide a DOI for the dataset which can be referenced in the thesis.

The duration of data preservation should be consistent with any data policies listed on q2. The [University of Southampton Research Data Policy](#) stipulates that significant data are preserved for a minimum of 10 years from the end of the project or publication date. Note: preserving data in a data archive is best practice and does not automatically mean that the data are openly available (see q8 below).

For software, the appropriateness of the answer should be judged on how well it matches the nature of the software, script or model being generated. Archiving a snapshot of the code alongside the data is a good option for scripts that are peculiar to the project, but not for software applications intended for many users. For the latter, a source code repository service would be more appropriate.

8. What are your plans for sharing the data/research materials after the submission of your thesis?

This section should state that:

- Data underpinning publications will be made openly available at time of publication; or
- Data underpinning publications will be made openly available after an embargo period; or
- Provide justification for data not being made openly available and the reasons for the need for any restrictions to be placed on data sharing (short, long-term or permanent). For more information see: [researchdata sharing](#)

UKRI funders, some other funders (e.g. Wellcome Trust, CRUK, Royal Society), and the University of Southampton Research Data Policy have requirements for data to be shared openly with as few restrictions as possible, but with caveats for situations where open sharing of data would breach legislative, regulatory, contractual, ethical obligations or IP exploitation.

What happens next

The DMP should be uploaded into PGR tracker. It has to be updated and reviewed by the supervisory team as part of every progression review and before the final submission of the thesis. This should make the completion of the Permission to Deposit Thesis form more straightforward. Dates of changes should be listed on the revision table on the first page of the document and revised copies should be uploaded to PGR Tracker.

Further help

The Research Data service can be contacted on researchdata@soton.ac.uk

