### About your Research

**PhD title:**

**Student name:**

**Supervisor(s):**

**Ethics No. (if appropriate):**

### About this plan

<table>
<thead>
<tr>
<th>Date of plan:</th>
<th>Frequency of reviews</th>
<th>12m / 6m / 3m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of next review:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Agreed actions to help you implement the plan:**
*For example training required on using bibliographic software (e.g. Endnote), Nvivo, writing R-scripts or anonymising data effectively*

**Agreed equipment and/or resources required:**
*For example space to be allocated in secure filing cabinet/Research filestore*

**Further information (as appropriate):**

### Version Table

<table>
<thead>
<tr>
<th>Version</th>
<th>Changes made</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Project Description:

2. What policies will apply to your research?

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

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

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

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

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

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

The University of Southampton Library has developed this Doctoral Research Data Management Plan and guidance notes based on material adapted from the Australian National Data Service, Sheffield Hallam University, the Open University and the universities of Bath and Newcastle.
Explanatory Notes

What is data?

The term data can be misleading as, in this context, it does not mean Big Data, electronic data or spreadsheets. It means, as the plural of the Latin datum, pieces of information whatever format they are in. In other words, your research materials which you use to answer your research questions and draw your conclusions.

For an historian, these may chiefly be a bibliography or primary and secondary sources and research notes based on those sources, with some additional working copy images of archival material. For a medic, they could be slides of tissue samples, experimental results and patient histories.

What are data management plans?

A data management plan (DMP) is a document that describes:

- What data/research materials will be created
- What policies will apply to the data/research materials
- Who will own and have access to the data/research materials
- What data management practices will be used
- What facilities and equipment will be required
- Who will be responsible for each of these activities?

What do I do with this plan?

You should discuss your plan with your supervisory team and it should be uploaded into PGRTracker as part of your progression review documentation. A DMP is a living document so you should revisit it as often as you feel is necessary but at least by every progression review to make sure it is still relevant. Any training or equipment needs which are highlighted in the DMP should be fed into your regular Academic Needs Analysis.

Why do I need a data management plan?

The carrot: improvements to efficiency, protection, quality and exposure.

Data management in some form is an unavoidable consequence of working with data. Typically data management is done at the last minute and using the first method that comes to mind. This approach is usually time-consuming and error-prone. Taking time at the start of a research project to put in place robust, easy-to-use data management procedures will usually pay off several times over in the later stages of the project. Inadequate data management can also lead to catastrophes like the loss of data or the violation of people’s privacy.
The stick: basic data management is required by the University as part of its Data Management Policy, [http://www.calendar.soton.ac.uk/sectionIV/research-data-management.html](http://www.calendar.soton.ac.uk/sectionIV/research-data-management.html), and also by many of the major funders of PhD studentships.

### What does a data management plan need to cover?

The following list of topics can be treated as a check-list:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backups</td>
<td>This is probably the single most important item on this list. You must have a credible backup strategy of regular backups, and of course you must then follow it. Consider including an off-site backup so that your data will not be lost if your building burns down. Consider an automated backup process.</td>
</tr>
<tr>
<td>Survey of existing data</td>
<td>What existing data will need to be managed?</td>
</tr>
<tr>
<td>Data to be created</td>
<td>What data will your project create?</td>
</tr>
<tr>
<td>Data owners &amp; stakeholders</td>
<td>Who will own the data created, and who would be interested in it?</td>
</tr>
<tr>
<td>File formats</td>
<td>What file formats will you use for your data?</td>
</tr>
<tr>
<td>Metadata</td>
<td>What metadata will you keep? What format or standard will you follow?</td>
</tr>
<tr>
<td>Access and security</td>
<td>Who will have access to your data? If the data is sensitive, how will you protect it from unauthorised access?</td>
</tr>
<tr>
<td>Data organisation</td>
<td>How will you name your data files? How will you organise your data into folders? How will you manage transfers and synchronisation of data between different machines? How will you manage collaborative writing with your colleagues? How will you keep track of the different versions of your data files and documents?</td>
</tr>
<tr>
<td>Storage</td>
<td>Where will your data be stored? Who will pay for the hardware? Who will manage it?</td>
</tr>
<tr>
<td>Bibliography management</td>
<td>What bibliography management tools will you use? How will you share references with the other members of your group/supervisor?</td>
</tr>
<tr>
<td>Data sharing, publishing and archiving</td>
<td>What data will you share with others? What license will you apply?</td>
</tr>
<tr>
<td>Destruction</td>
<td>What data will you destroy? When? How?</td>
</tr>
<tr>
<td>Responsibilities</td>
<td>Who will be responsible for each of the items in this plan?</td>
</tr>
<tr>
<td>Anything else</td>
<td>Don't restrict yourself to the items above. Stop and think. What is missing from this list? If you think of something, please let us know so that we can update this information.</td>
</tr>
</tbody>
</table>
Detailed Guidance Notes

Also see the Data Plan for your PhD webpages: http://library.soton.ac.uk/researchdata/phd

For additional support, email researchdata@soton.ac.uk

1. Project Description

Provide two or three sentences summarising your project's research questions and data needs.

2. What policies will apply?

University policies that might be relevant to your project are listed below:

- Open Access Policy: http://www.calendar.soton.ac.uk/sectionIV/open-access.html
- Ethics Policies: https://www.southampton.ac.uk/about/governance/regulations-policies-guidelines.page#research_%26_enterprise_policies
  https://intranet.soton.ac.uk/sites/researcherportal/Lists/Services1/testing.aspx?ID=285&RootFolder=%2A

Working with personal or sensitive data?

If humans are involved in your research, you will need to take measures to comply with the Data Protection Act 2018 and GDPR. Consult the University's guidance to find out what this will mean in practice for your research: https://www.southampton.ac.uk/legalservices/policy-and-guidance.page

3. What data will you collect or create?

- What physical data will you study? (e.g. artefacts, samples, paper archives, etc.)
  And what digital data will be derived from these? (e.g. field-notes, images, measurements, spreadsheets, survey data, etc.)
- How will the data be collected? Is it gathered from experiments? From the literature? What instruments? How about observations or photos?
- Will you be using secondary data?
- Could the data be considered personal, sensitive or commercial data?
- Describe the methods/standards for data creation. What quality assurance processes will you adopt (e.g. calibration, data entry validation, representation with controlled vocabularies)
- What file formats and software will you use? Do your chosen formats and software enable sharing and long-term sustainability of data, such as open standards and open source software?
- Consider how many individual files you expect to make, anticipated file sizes, and total storage volume.
- Frequency of new data - how often will you get new data and over what time period?
  Continuously or just from discrete experiments? How many experiments per week? How will this change over time?
Examples:

I record interviews with subjects using a digital audio recorder, then transcribe them into text.

I test my catalyst under a number of conditions, then submit samples of the products to analysis facilities.

I generate data using model code that I've written, then process it in various ways to produce visualisations.

I combine existing data from a number of sources [e.g....] and reanalyse them to derive new conclusions.

All of my data, part from my literature search, will come from a single 3-month field trip to various archives in France in my second year.

I expect to run two or three experiments each week through my second year and much of my third year – about 100 in total.

4. How will your data be documented and described

Think about what contextual information is required to make the data understandable to others (and yourself in three years’ time!).

- Has a file naming convention and directory structure been agreed? (e.g. date created/date amended/version no.)
- What information on the data collection methods and context (documentation and ‘metadata’) will be recorded for each data type/set?
- Where will the metadata for each data type/set be located? (e.g. within the data file and/or as separate metadata text document, and/or in method chapter/appendices in the thesis)
- How will you tell different versions of the data do you create apart? For example, versions of data files
- Do you update or add data to existing files?
- Describe the system to name and structure any electronic files. Are there any set or recommended standards in your discipline?

Examples:

I use the structure <archive collection>/<mss no> for transcripts, notes on documents and working copy images.
Filenames are suffixed with transcript, notes or page nos as appropriate.

Each filename starts with the date on which the data was collected using the format YYYYMMDD. As I survey new cohorts, data is appended to the dataset and saved as a new file.

There is only ever one version of each data file — new experiments create new data, which is stored in a new set of files with machine generated filenames. I keep a register of filenames and the experiments they relate to.

Each time I run a new version of my model, intermediate files are written over, but the final results are saved as a new file.

Weekly check that files on the R: drive are still usable.

Working data is backed up on the UoS Research Filestore. I will make sure I copy the latest versions of my working files there each day.

I regularly scan my paper-notebook and store digital copies on the University storage.
5. How will you deal with any ethical and copyright issues?

- Who else has a right to see or use this data, even before you share it? If your data is personal, sensitive or commercial how will you share safely, including plans to anonymise your data?
- Do you need to anonymise data during research or when preparing for sharing, and how will you do this?
- Have you established who owns the copyright in your data?
- If you are re-using someone else’s data, are there any restrictions on their re-use?
- Could the data be considered high value and/or vulnerable? For example, is your data likely to attract “hacktivists”? How could this be mitigated?
- How will you destroy any personal, sensitive or commercial data identified above?

Examples:

- I will share my data with my research group/supervisor using a shared folder. Due to the sensitive nature of my data I will encrypt my data and send via Dropoff (dropoff.soton.ac.uk) to my collaborators.
- My data will be pseudo-anonymised prior to sharing, with files encrypted.
- My data is of high value and may be subject to commercial sabotage, I will check for advice in the Information Security Best Practice: https://intranet.soton.ac.uk/sites/gdpr/Pages/Information-security-best-practice.aspx and contact Information Security team in iSolutions for guidance.
- My paper based notes from interviews will be shredded using confidential waste. My electronic files will be overwritten multiple times using specialist software, for example Eraser https://eraser.heidi.ie/

6. How will your data be stored and backed up?

- Do you know the backup procedures of the storage space?
- Quantity of data (Megabytes, Gigabytes, Terabytes, other forms of storage)
- Where will the data be stored? For electronic data there should be 3 places, University storage should be one of the locations.
- If keeping your own copy of the data are there security considerations, e.g. encrypted flash drive? How will you know which is the master copy?
- How much have you got so far? Try to estimate how this will grow for the rest of the project
- Describe the regime for backing up the data.
- Describe the procedure to be used to ensure files can be restored from the backups.

Example:

*Each experiment produces about 50MB of data, so over the course of my PhD I expect this to add up to about 5GB, plus two drawers of a standard filing cabinet*

*My primary copy of my bibliography is on my laptop. I make weekly back-ups of it to my University filestore H: drive every Friday afternoon*
7. What are the plans for the long-term preservation of data supporting your research?

- What data/research material should be kept beyond the end of the project? Refer to any ethics approval documentation if appropriate.
- What data/research material should be destroyed for contractual, legal or other purposes?
- How long will you preserve your data for?
- Where will you preserve your data? In the UK Data Archive? In the UoS Institutional Repository?
- How will you prepare and document the data for preservation?
- What file formats can you export to for long term preservation?

Examples:

I am responsible for archiving data, and the archive service will maintain it for a minimum of 10 years as per the University RDM Policy.

The data is part of a larger project and will be archived with the project; my supervisor will deal with this.

All data, both raw and processed will be retained. Spreadsheets will be saved as csv files.

Only simulation code and input parameters will be kept.

Transcripts of all interviews, but not recordings. Personal data and anonymization key will need to be destroyed securely at the end of the project.

8. What are your plans for data sharing after submission of your thesis?

- Will any of the digital data supporting the thesis (e.g. organised project archive folders with images, drawings, spreadsheets, databases, etc.) be made available to others via a repository?
- Are there funding body/institutional requirements for the re-use of, or open-access to, the data?
- What are your supervisor’s thoughts on sharing ‘their’ research data, if on a project team?
- With whom will you share your data and under what conditions? Should anybody be able to download the data, or is there a need for access restrictions (e.g. an embargo period, or making data available on request only)?
- Who, if any, are the anticipated future users of any digital data/resources from the research, e.g. yourself, project partners, future students, peer researchers, the public?
- Where will the data be archived?
- Who will create and maintain the archive of data?

Examples

Tables for household income and relative market prices of goods in my thesis will be made available as spreadsheets. My bibliography will be made available as a csv file so it can be reused by other scholars. These will be made publicly available in the institutional repository and linked to my ethesis. The working copy images of archival material cannot be shared due to copyright restrictions by the various archives, however I will upload the full transcripts of those documents which I quote extensively in my thesis if allowed by the relevant archives.

All my experimental data will be made available on the institutional repository, accompanied by a readme file describing the data and the data linked back to the relevant part of my thesis. The data will only be made available after a three year embargo period as I plan to publish further articles from my thesis.