

Research data management
planning:
An introduction for
researchers

How to Use this User Guide

This handbook accompanies the taught sessions for the course.

Later, you may attend follow-up sessions at ITLP called Computer8, where you can continue work on the exercises, with some support from IT teachers.

Files Used

Shotton-Twenty_Questions_for_Research_Data_Management.docx

Post-Graduate_DMP_Form.docx

Revision Information

Version	Date	Author	Changes made
1.0	October 2014	Meriel Patrick	Created
1.1	February 2015	Meriel Patrick	Minor updates

Copyright



The IT Services Research Support team makes this document and the accompanying PowerPoint presentation available under a Creative Commons licence: Attribution, Non-Commercial, No Derivatives. Individual resources are subject to their own licencing conditions as listed.

The Oxford University logo and crest is copyright of Oxford University and may only be used by Oxford University members in accordance with the University's branding guidelines.

Contents

1 Introduction	1
1.1. What is research data management?	1
1.2. What is a data management plan?	1
1.3. Why make a data management plan?	1
1.4. When should a DMP be created?	2
2 Data management plan content.....	3
2.1. Describing the project and data	3
2.2. Handling data during the project.....	3
2.2.1. Oxford provision and services.....	4
2.3. Documentation and metadata	4
2.4. Long-term planning.....	5
2.4.1. Repositories and archives	5
2.5. Meeting requirements	6
2.5.1. Legal and ethical requirements	6
2.5.2. Funder and institutional requirements	6
2.6. Practical planning.....	6
3 Further resources and guidance	7
3.1. Research Data Oxford website	7
3.2. Digital Curation Centre.....	7
3.3. UK Data Archive	7
3.4. University of Bristol resources	7
3.5. Research Support team, IT Services.....	7
4 Data management planning tools.....	9
4.1. DMP Online.....	9
4.2. Twenty Questions for Research Data Management.....	9
4.3. Other templates	9
Appendix: Student Exercise.....	10

1 Introduction

1.1. What is research data management?

Research data management is an umbrella term – it covers both dealing with data on a day-to-day basis during the lifetime of a research project (organizing it, structuring it, choosing appropriate software or other technologies to manage it with, storing it, backing it up, and so forth) and longer term issues (sharing and preservation, for example).

1.2. What is a data management plan?

A data management plan (or DMP) is a document which outlines how data will be managed over the course of a project. It details plans and expectations for the data which will be used. Key topics that would normally be covered include:

- Nature of data and its creation or acquisition
- How it will be documented
- Storage and security
- Legal and ethical issues
- Preservation and sharing

1.3. Why make a data management plan?

Some funding bodies now require researchers to provide a data management plan as part of the application process. This reflects the increased interest in research data taken by funders over the last few years – data is increasingly being recognized as a valuable research output in its own right.

Even if your funder doesn't require a plan, there are still good reasons to make one. It's a chance to anticipate possible problems before they occur – and look for solutions in good time.

Making a plan will often save time and reduce stress later in the project. Many aspects of data management are straightforward if they're planned for from the beginning, but much harder to do retrospectively. Planning ahead can bring particular benefits when it comes to preparing data for sharing. For example, documenting what's happened to data can be done quickly and easily if good recording processes are built into the research methodology, but trying to unpick what's been done later on may be almost impossible.

Dwight D. Eisenhower, who served as a general in World War II and then went on to become president of the USA, is on record as having said 'In preparing for battle, I have always found that plans are useless but planning is indispensable.' While a research project isn't (or shouldn't be) a battle, President Eisenhower's words nevertheless have some relevance in this context. It is almost inevitable that unexpected events will arise – it's very rare that everything goes exactly as anticipated. But although this means you may often have to adapt your plan to take account of new developments, this makes having created a plan in the first place more essential, not less. If you've thought through all the relevant issues, you're less likely to be taken by surprise – and you'll be better placed to respond when the unexpected does crop up.

1.4. When should a DMP be created?

The earlier in a project a plan is created, the better. However, it's never too late: even if a project is starting to draw to a close, it may still be worth investing some time in planning for the life of the data beyond the end of the project.

Plans are often created when applying for funding or setting up a project. It's common to create an initial outline plan at the funding bid or initial idea stage, and fill this out later when the project is getting underway.

In fact, it's best to think of planning as a process, rather than a one off task. Things will change and new circumstances will arise, so it's good practice to revisit and revise plans regularly. A schedule for doing this can be included in the initial version of the plan – you might, for example, plan to see if the plan needs updating every six months.

2 Data management plan content

There are various ways of structuring a data management plan, but a typical plan would contain these elements:

- Description of the project and the data
- Plans for handling the data during the project
- Plans for documentation and metadata
- Long-term plans for the data
- Consideration of requirements that need to be met (legal, ethical, institutional, funder)
- Practical planning to ensure the plan can be implemented

2.1. Describing the project and data

It's helpful to begin a plan with a brief description of the project's subject matter and overall goals, to set things in context. It's also worth recording who'll be working on the project: if it's a team endeavour, each person's role should also be noted.

The main content of this section of the plan will be a description of the data it is anticipated will be used during the project. This would include:

- Details of the type of data – will it be textual, numerical, images, videos, or something else? Many projects will use a combination of types of data
- An estimate of the quantity of data, in terms of number of files or records, and/or their size. It may be hard to give an accurate figure at this stage, but having some idea of how much data is expected (even if this is just an indication of whether it should be thought of in terms of megabytes, gigabytes, or terabytes) makes it easier to select appropriate storage solutions.
- Information about the file formats that will be used.

Information should also be given about where the data will come from. Will it be collected or generated during the project? If so, give a brief description of the methods that will be used. If data from earlier projects, or third party data from an external source such as an archive or commercial supplier is being used, note this and the source(s).

2.2. Handling data during the project

This section of the plan should cover a range of practical issues to do with how the data will be dealt with during the active phase of research – while data is being collected, manipulated, and analysed.

Key considerations here include:

- Where will the data be stored? If the size of the data isn't vast, it can be tempting simply to keep everything on a personal device such as a laptop. But this isn't always the best option: other solutions may offer better security, and/or make it easier to share data with colleagues.
- What software will be used to store and analyse the data?

- How will the data be backed up? What plans are in place to ensure that back-ups actually happen? (For example, can backing-up be automated?)
- Does data need to be shared within the project team? If so, how will this be achieved?
- If any of the data is confidential or otherwise sensitive, how will it be kept secure?
- How will versioning be handled – that is, how will the project keep track of subsequent versions of the same dataset, and ensure that the appropriate version (usually the most recent one) is being used? This is particularly important if multiple people will be working on the same dataset.

2.2.1. Oxford provision and services

Oxford provides various resources to help researchers store their data securely.

- You may be entitled to storage space on a departmental server. This can be a good option, as it often provides an easy way of sharing data within a project team, and departmental servers are usually automatically backed up.
 - Ask your department's IT officer if this option is available to you.
- Nexus SharePoint can also be used for sharing files.
 - See <http://help.it.ox.ac.uk/nexus/sharepoint/>
- If your department can't provide what you need, IT Services' NSMS can provide server space –although this is a chargeable service.
 - See <http://www.it.ox.ac.uk/nsms/>
- The recently launched ORDS (Online Research Database Service) allows creation of online relational databases. You can import an existing data collection, or create a database from scratch. Online hosting makes it easy to share data with collaborators.
 - See <http://ords.ox.ac.uk/>
- The HFS back-up service is free to staff and postgraduate students. This can be set to automatically back up computers connected to the University network.
 - See <http://help.it.ox.ac.uk/hfs/>
- IT Services' InfoSec can advise on security.
 - See <http://help.it.ox.ac.uk/service/information-security>

2.3. Documentation and metadata

Another important aspect of day-to-day data handling is documentation – that is, ensuring that all relevant contextual information needed to permit proper interpretation of the data is recorded. Documentation can be thought of as a user's guide to the data. This might include describing methods, recording how data has been manipulated, and ensuring that acronyms, abbreviations, and coding are adequately explained.

In this section of the plan, you should state how the data will be documented and described. It's also worth thinking about whether practice can be standardized, particularly for projects where a team of people are involved. Consistent

conventions for naming files, or for recording what has been done to or with data, can make it much easier to navigate a large body of information. It's best if these conventions can be in place right from the start of the project – although sometimes they may need to evolve as the work progresses.

Metadata is literally data about data, and is a term often used to refer to a particular type of documentation which has a defined structure and is often designed to be machine readable – it may be used, for example, to create a catalogue record for a dataset deposited in an archive. The information that's useful varies between subject areas, and hence many academic disciplines have metadata standards, or recommended formats in which data about datasets should be presented. It's worth being aware of these in advance, so you can ensure you're capturing all the relevant information. The Digital Curation Centre has collected information about disciplinary standards on their website – see <http://www.dcc.ac.uk/resources/metadata-standards>

2.4. Long-term planning

When you're in the planning stages, the end of a project can seem a long way off. However, it's well worth giving some thought to longer-term issues early on – in particular, what will happen to the data at the end of the project.

Data is increasingly being recognized as a valuable research output in its own right, and consequently researchers are being encouraged to ensure that data is preserved, and where possible, made available for use by others.

Presenting data in a format that's suitable for re-use is likely to require some work, but effort can be minimized if processes for doing this are built into the project's workflows from the beginning.

This section of a data management plan should explain the long-term plans for the data used during the project, and consider what needs to be done during the rest of the project to facilitate these plans.

2.4.1. Repositories and archives

One of the best ways of making data available for re-use is to deposit a copy in a specialist data repository or archive. These are designed to provide a secure long-term home for data, and mean that once data has been deposited, the researcher no longer needs to worry about curation. It's usually possible to embargo datasets for a fixed period (to allow the creators of the data time to publish their findings, for example), or to impose restrictions on re-use.

Many national disciplinary archives exist. Two searchable catalogues of these are available: Databib can be found at <http://databib.org/>, and Re3Data.org at <http://www.re3data.org/>. It can be worth identifying a suitable repository well ahead of time: this allows you, for example, to follow any guidelines they have about how data should be presented.

For datasets which need to be preserved but which don't have another obvious home, Oxford has its own institutional data archive, ORA-Data. The pilot version of this was launched in late 2014, with a view to moving towards a full service by May 2014. ORA-Data is designed to work in parallel with the existing archive for textual research outputs, which will become known as ORA-Publications. For more information, see <http://www.bodleian.ox.ac.uk/bdlss/digital-services/data-archiving>.

2.5. Meeting requirements

Measures to ensure some requirements have been met may have been included in earlier sections, but it's worth taking a special look at this to ensure everything has been covered.

2.5.1. Legal and ethical requirements

A key issue to clarify is who owns the data used during the project. If data has come from multiple sources, or been worked on by different groups, the rights situation may be complex. You may need to seek specialist help – contact the Research Support team at IT Services (researchsupport@it.ox.ac.uk) for advice.

If you're using third party data, or confidential information, there may be restrictions on what can be done with the data. It's important that these are clearly documented, and a plan made for ensuring that any restrictions aren't violated. If data is sensitive, this will include a plan for ensuring an adequate level of security is maintained.

If you're hoping to (or are required to) share data at the end of a project, there may be steps it's necessary to take earlier on to allow this to happen – you may need to ask research subjects for appropriate consent, for example, or to make a plan for anonymizing data.

If data sensitivity issues mean that some data will need to be destroyed after use, this should also be noted, and a plan for doing so should be put in place.

2.5.2. Funder and institutional requirements

Many funding bodies now impose requirements relating to data – for example, researchers may be asked to publish their data at the end of a project, unless there's a good reason not to do this. It's worth checking your funder's requirements early on, so you can ensure that adequate plans for meeting them have been made.

Researchers should also be aware of the University of Oxford Policy on the Management of Research Data and Records. This makes a number of stipulations about data integrity – that data should be complete, accurate, retrievable, and secure, for example. It also states that data should be retained for *at least* three years after the end of a research project (longer if required by funders), and made available for use by other researchers wherever possible.

A copy of the policy can be seen on the Research Data Oxford website: <http://researchdata.ox.ac.uk/>

2.6. Practical planning

Finally, a data management plan should consider how the plan itself will be implemented. It's helpful to agree responsibilities for putting each element of the plan into action, and to ascertain at an early stage whether any additional resources (including extra time) will be needed.

As noted above, planning should not be viewed as a one-off event, but rather as a process that continues throughout the research project. It's therefore good practice to consider how adherence to the plan be evaluated, and when the plan will be revisited. For example, you might plan to assess the project's data management practice every six months, and to use that as an opportunity to revise the plan to reflect current circumstances.

3 Further resources and guidance

3.1. Research Data Oxford website

Research Data Oxford is a central advisory website for members of the University of Oxford. It covers all aspects of research data management, including giving an overview of funder requirements. The University of Oxford Policy on the Management of Research Data and Records is also available.

If you have any questions, you can contact the University's cross-departmental advisory group by emailing researchdata@ox.ac.uk

The Research Data Oxford website can be found at <http://researchdata.ox.ac.uk/>

3.2. Digital Curation Centre

The DCC is a national service providing advice and resources. Although their focus is on data preservation, they cover the whole research process, and have a wealth of resources relating to data management planning.

These include the online guide 'How to develop a data management and sharing plan', a checklist for DMPs, example plans, and a summary of funders' requirements.

The DCC also provides the web-based DMP Online tool – see below for more details.

The DCC's website can be found at <http://www.dcc.ac.uk/>

3.3. UK Data Archive

As well as being the UK's largest collection of digital social sciences and humanities research data, the UK Data Archive provides extensive advice on how to deal with research data, covering storage, documentation, planning for sharing, and a range of other topics.

One particularly useful resource for those working with human subjects is the section on consent and ethics. This includes sample consent forms, which provide some pointers on how to provide subjects with an appropriate level of confidentiality, while also making as much data as possible available for re-use by other researchers.

The UK Data Archive's website can be found at <http://www.data-archive.ac.uk/>

3.4. University of Bristol resources

The University of Bristol offers a useful set of online guides to writing a data management plan for each of the RCUK funding bodies, plus some others. Their site also includes more general advice about sharing, storage, anonymization, etc. (But do bear in mind that it's aimed at Bristol researchers, so may contain references to Bristol-specific services.)

3.5. Research Support team, IT Services

The Research Support team at IT Services is available to help Oxford researchers with all technical aspects of research projects, including data management. They can provide help writing a plan, or give feedback on a draft, and are also available to advise on selecting appropriate software and storage solutions, and on data modelling.

Research data management planning: An introduction for researchers

You can contact the Research Support team by emailing researchsupport@it.ox.ac.uk, or visit their website at <http://research.it.ox.ac.uk/>

4 Data management planning tools

4.1. DMP Online

The Digital Curation Centre offers a free online data management planning tool called DMP Online. This offers a variety of templates, to meet the requirements of various funding bodies, plus guidance on how to complete them. Plans can be revised and expanded over time, and can also be shared with colleagues or exported for use elsewhere.

DMP Online can be found at <https://dmponline.dcc.ac.uk/>

4.2. Twenty Questions for Research Data Management

Another resource for building a DMP is 'Twenty Questions for Research Data Management', created by Oxford academic David Shotton. The questions are accompanied by sample answers to aid completion, and answering all the questions is sufficient to create a basic data management plan.

A hard copy is provided in these notes, and an electronic copy is available on the H: drive. You can also download a copy from

<http://datamanagementplanning.wordpress.com/2012/03/07/twenty-questions-for-research-data-management/>

4.3. Other templates

Plenty of other templates also exist. A simple one designed for use by graduate students is also included with these notes, and available on the H: drive.

Appendix: Student Exercise

Exercise 1 Drafting a data management plan

- *Start the process of drafting a data management plan for your own research*

Using either DMP Online (<https://dmponline.dcc.ac.uk/>) or one or the other DMP templates, start the process of drafting a data management plan for your own research.

You may find there are questions you can't answer at this stage. If this happens, make a note of what you need to find out, or what decisions need to be made.

Remember to save a copy of your work so you can continue the drafting process later.