Writing a Data Management Plan for a Grant Application

Table of Contents

Introduction

This guide to writing a data management plan for a grant application focuses on general principles, and provides information about University policies and services. Separate guides on writing data management plans (DMPs) to meet the requirements of some of the main funders are available for download here.

Information about funders that have data management policies/requirements and availability of UoR guidance can be found in the Appendix.

Where a DMP is required, the lead applicant must send a draft of the DMP to the Research Data Manager for review prior to submission. The draft DMP should be accompanied by the latest version of the application proposal, and must be sent to the Research Data Manager no later than five working days before the internal or funder submission deadline.

Applicants are encouraged to contact the Research Data Manager earlier in the application development process, where they require help in completing the DMP or need to specify costs for data storage, computing or archiving.

Contact: Research Data Manager: researchdata@reading.ac.uk / 6161.

Why is a DMP required?

Primary data collected or created in support of project findings are integral to the validation of those findings, and may have long-term value to future researchers and other users (e.g. policy-makers, industry users, teachers and students). Many funders expect their researchers to manage the data they collect responsibly, and to ensure these data are preserved and remain as widely accessible as possible over the long term. To achieve this end, data supporting project findings should be archived in a suitable public data repository and made openly accessible, unless there is a valid legal, ethical or commercial reason for restricting access to them.

These requirements are expressed in the research data policies of the Research Councils, the European Commission, the Royal Society and the Wellcome Trust, as well as many other public and charitable funders around the world. They are also stated in the University's Research Data Management Policy. For more information about funders' policies see guidance here.

Not all data have to be preserved. The requirement to preserve data may not apply where the data do not have significant value (for example, test data) or where the cost of archiving is in excess of any likely benefit (as may be the case with certain model output data, for example).

What is a Data Management Plan?

A Data Management Plan is any section of a grant application where the applicant is required to address the collection, management, preservation and sharing of research data. The DMP may be a separate attachment in addition to the main application (as is

the case with most Research Councils) or a section of the application form (as with Royal Society grants).

Most funders have their own DMP templates that applicants are expected to use, or specify a set of points that the DMP must cover. DMPs can vary in length, from a few lines on deposit of datasets for the British Academy to 3 pages for ESRC, and the specific questions to be answered will vary. Shorter DMPs tend to focus on identifying data outputs and the means by which they will be preserved and shared. Longer DMPs may also ask about aspects of data management during the project, and legal and ethical matters.

This is a list of common components that may be covered in a DMP:

- The data outputs that will be generated by the research.
- Procedures that will be used for data collection, processing, documentation, and quality control.
- How data will be stored and backed up and shared between members of the team during the project.
- How legal and ethical aspects of data management will be handled, including compliance with data protection and research ethics obligations.
- Who will have ownership of the data/how intellectual property in the data will be managed.
- Where data will be archived for long-term preservation and access.
- Whether any restrictions will placed on access to the data for legal, ethical or commercial reasons.
- When data will be made accessible.

A DMP can also help applicants to identify where there may be costs for storage of data during the project, for computing resources required to generate/process data, and for archiving of data on project completion. These costs may need to be included in the project budget. For more information, see the section on Costs, below.

Completing the DMP

Guidance on addressing the components of a DMP follows.

Data outputs

Data outputs that will be generated by the project should be succinctly described. Realistic indications of data volume may be particularly important where the volume is likely to be substantial (as a rough guide, greater than 100 GB), and there may be costs for data storage.

You should briefly itemise and describe the types of data you will collect or create in order to answer your research questions. You need only describe the categories of data or the key variables, sufficient to identify their essential characteristics, e.g. telephone interviews, EEG brain activity measurements, weather system model outputs.

Any code you may write to process or analyse data should also be described where possible. If these will be required to replicate your final results, they should be preserved and made accessible to others as well as your data.

For each data type, you should be able to say what physical and digital file formats data will be stored in. Bear in mind that data may be collected, processed, analysed and preserved in different formats. For long-term preservation of data collected in proprietary file formats, e.g. specific to a data collection instrument, such as a digital camera or an NMR spectrometer, or to proprietary software, such as MATLAB, you should plan to save data to an open format or one accessible without a software licence where possible, such as CSV for tabular data, or JPEG for images. Widely-used formats are acceptable, e.g. MS Word. The UK Data Service provides guidance on recommended file formats for commonly-used file types.

For each data type, try to quantify the amount of data you expect to collect, e.g. in numbers of experiments/observations/interviews, duration of collection period and frequency of data collection, or overall numbers of bytes (MB, GB, TB). You may be able estimate total data volume by calculating the numbers of files of each type you intend to collect, and the average size of each file.

Data collection, processing, documentation and quality control

Some funders will ask about methodologies for data collection and management, how data will be documented, and procedures for data validation and quality control.

You may want to provide information about proposed instruments and methods for data collection. Instruments may include hardware, software and paper-based instruments, e.g. data collection forms, lab notebooks. If you will be using any experimental facilities, e.g. the ISIS neutron and muon source, or research infrastructure, such as the NERC ARCHER supercomputing service, this should be indicated.

You should plan to use standard procedures wherever possible in your research methods, and to use standard data and metadata formats to record the information you collect. Standards promote reproducibility of research, and understanding and re-use of data.

Particularly in laboratory-based work there may be standards for recording information about experiments, such as the Minimum Information About a Microarray Experiment (MIAME) standard promulgated by the Functional Genomics Data Society. Experimental protocols should be recorded in detail - tools such as https://www.protocols.io/ allow protocols to be developed, shared, annotated and linked to published findings under version control.

If you will be using specific data formats or metadata standards to describe particular data types, e.g. to comply with data repository requirements, you may identify these and provide relevant information. For example, the European Nucleotide Archive specifies data formats for submission. Search for relevant standards at FAIRsharing. For more information about Metadata Standards, see guidance from the Research Data Alliance.

Quality control should be considered at the time of data collection, data entry, digitisation or data checking, to ensure accuracy and consistency of data. Procedures might include: documenting the calibration of instruments, the collection of duplicate samples, data entry methods, data entry validation techniques, methods of transcription. Ensure any scripts you write to execute workflows are preserved and documented.

Storage and back-up

Data collected/held at the University should be stored using University-managed infrastructure, which will provide data security, replication in separate data centres, automated backup and file recovery. For the different options available, and information about costs, please read the guidance here.

Sensitive/confidential data can also be stored in these locations; if such data are stored or shared using other devices or cloud services, this should be in accordance with the University Data Protection, Remote Working and Encryption Policies.

Data collected in the field should be stored securely and backed up using local devices and transferred at the earliest opportunity to the primary storage location.

If you have computing-intensive requirements, custom specifications of CPU, memory, storage and GPU can be purchased from the University on a pro rata basis. Information is available in the Academic Computing Team website.

If data are acquired using specialist infrastructure, such as the ISIS neutron and muon source, or the JASMIN supercomputing environment, raw data may be stored in the facility infrastructure, and data copied or extracted locally as required.

You may need to address storage of non-digital data, e.g. signed consent forms, in appropriate secure environments.

Research ethics and data protection

You have an ethical obligation to protect the confidentiality of personal information provided to you by research participants, and you must also comply with data protection law if you collect and process personal data. Where personal data are processed in jurisdictions outside the UK, the standards prescribed by UK data protection law must still be observed.

Any research involving human subjects will need to receive approval from your School's or the University's Research Ethics Committee. Guidance can be found here.

In most cases data can be shared openly if they are anonymised and appropriate consent procedures have been followed. The UK Data Service provides excellent guidance on consent and anonymisation, and has sample information sheets and consent forms.

If you will be processing personal data in your research, you are advised to consult University guidance on Data Protection and Research. The Data Protection Checklist for researchers is a good starting point. You can also find here a sample information sheet and consent form.

Personal data is any information relating to an identified or identifiable natural person. These data enjoy statutory protection under the General Data Protection Regulation 2016 and the Data Protection Act 2018. Under this legislation any personal data collected by you must be processed fairly and lawfully. Among other things you will be required to issue a Privacy Notice to your research participants, which explains the purpose(s) for which the data are being collected, your lawful basis for processing the data, who the data will be disclosed to, and the rights of the individuals in respect of their personal data. For certain kinds of research, for example involving the processing of sensitive data or human genetic data, you will need to complete a Data Protection Impact Assessment under the advice of the University Information Management & Policy Services Officer.

You must ensure that personal data are kept secure and are not disclosed to unauthorised persons. You should use a locked storage container such as a filing cabinet in a locked office for paper-based personal data; for digital data, password-protected or, preferably, encrypted storage. This particularly applies in the case of special category sensitive personal data, which include information about an individual's: race; ethnic origin; politics; religion; trade union membership; genetics; biometrics (where used for ID purposes); health; sex life; or sexual orientation. Such personal data should be encrypted, and not stored or shared by means of cloud services other than a University OneDrive account, or transferred via unencrypted channels (e.g. via email). You can transfer data to a location on the University network using VPN, which provides an encrypted channel.

Working procedures should be designed to minimise the risk of inappropriate disclosure. Data can often be pseudonymised for purposes of processing and analysis, with the personally-identifying information and their linked IDs stored separately from the working dataset. When the study is complete and if there is no further need to link individuals to data, the linking key can be destroyed, so that the data become anonymised.

You can retain personal data on a continued basis for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. You do not need to commit to destroy personal data at a set time, but they should be managed under a retention schedule that specifies periodic reviews, so that they can be securely destroyed when no longer needed.

Intellectual property

In the absence of any contract stating otherwise, ownership of data created by employees of the University in the course of their employment will be vested in the University. It is standard in collaborations for each institution to own IPR in the data it has created. Where data are jointly created, IPR will be shared.

Where secondary sources will be used, you should investigate existing IPR in these sources to ensure they will not inhibit use of the data for the purposes of the project. If terms of use or licence conditions are not clearly stated, these should be investigated with the data provider. Where you plan to produce derived outputs that may incorporate

or build on existing resources, you should seek permission to share these where possible.

Where copyright materials are contributed by research participants (e.g. photographs) you should ask them to transfer copyright in the materials to the University or to grant the University a licence to use and distribute the materials.

Archiving

Researchers should where possible use UKRI-funded or other data type-specific repositories to preserve and enable access to data. UKRI data repositories include the NERC data centres, the ESRC's UK Data Service ReShare repository, and the Archaeology Data Service, funded by AHRC and NERC. BBSRC contributes to a number of international bioscience data sharing resources, including the molecular biology databases of the European Bioinformatics Institute. The Wellcome Trust also maintains a useful list of approved data repositories. For a more general search, consult the data repository registries FAIRsharing or re3data.org.

Where no suitable external service exists, researchers can deposit data in the University's Research Data Archive. Research data in non-digital formats and digital data that cannot be made accessible or require controlled access should also be registered in the University Archive. The Archive can provide a mechanism to regulate access to controlled data under data sharing agreement where this is necessary.

Guidance on selecting a suitable data repository can be found here.

Most data repository services are free to use, but there are some exceptions, and high-volume dataset deposits may incur charges from some services, including the University's Research Data Archive. See the section on Costs for further information.

Remember that data will need to be accompanied by appropriate documentation, to enable others to understand and make use of them.

You should also archive any code written that is required to replicate results, e.g. data analysis scripts, model source code. For more information, see the section on Software and code, below.

Access to data

Data should be made accessible no later than publication of the main findings. If release of data may be delayed, for example, pending confirmation of IP protection, this should be stated in the DMP.

Access restrictions

If any legal, ethical or commercial reasons for restricting access to data are anticipated, this should be stated.

Where data are collected from human subjects, in most cases these can be anonymised for sharing. A valid reason for restricting access to such data would obtain only if it will not be possible to anonymise the data (e.g. biometric data may be intrinsically identifying) or if the risk of causing harm or distress by disclosure is significant.

Even data containing personal or confidential information may be shared under certain conditions, with appropriate consent. Some data repositories, e.g. UK Data Service ReShare repository and the European Genome-phenome Archive, can manage controlled access to sensitive/confidential data. The University Archive may also be able to offer a restricted access option. Contact the Research Data Manager for advice.

It is acceptable to restrict access to data if they are commercially confidential or there is a commercial pathway for the research, for example involving an identified industrial partner. If IP protection may be sought, it should be possible to release data once protection has been confirmed.

Costs

You may need to include costs for use of the Research Data Storage service or the Research Cloud computing platform. Information about these services and guidance on identifying a suitable specification for your requirements are available from the Academic Computing Team.

Archiving costs may also need to be included in your grant budget. Most data centres/data repositories do not charge for deposit of data. The Archaeology Data Service is one exception. The University Research Data Archive does not charge for deposits up to 20 GB; above this threshold deposits may be subject to a charge, which should be discussed with the Research Data Manager.

Software and code

If you create software or write code to script workflows for collecting, processing or analysing data, you should plan to preserve these and make them accessible as well as your data, as they will be integral to the computational reproducibility of your results.

There is a guide to publishing research software available for download here. This elaborates some of the information below.

In most cases short scripts and segments for code written to perform standard operations, e.g. for purposes of data processing, statistical analysis or data visualisation, can be archived (with any comments/documentation) alongside data, under the same licence as the dataset (for example, a Creative Commons Attribution licence).

Where the research software is more substantial or has been written in the context of an ongoing project or established community, a development-oriented approach to management and release of code may be appropriate, with source code released under an Open Source licence. This might be the case, for example, where code has been written to implement a simulation model, or where you have developed existing source code.

There are various flavours of Open Source licence, but they essentially come down to permissive (anyone can use the code for any purpose, including commercial), e.g. Apache, MIT; and copyleft (the code and any derived works must be licensed under the same terms – these prevent others from privatising the code and selling it), e.g. GNU

GPL. All licences feature an attribution requirement – i.e. the copyright owner must be acknowledged, and you can specify how you wish the item to be cited. There is a handy guide to licences at https://choosealicense.com/.

For managing and sharing code under ongoing development or with multiple contributors it is a good idea to use an online code repository platform, e.g. GitHub (widely-used) or GitLab, which is available as a University service. Both GitHub and GitLab are free to use (GitHub provides a free basic service, but you can pay for more features), and provide version control, code review, bug tracking, documentation, and other features. Repositories can be private or public, so that code can be maintained during a closed development phase and then released for open use at an appropriate stage.

Online code repositories are good solutions for ongoing software development and support. In order to ensure reproducibility you may also need to archive a specific version of the code used to generate results reported in a paper. If you archive a *version* of the code to a suitable repository, it will be preserved and can be assigned a DOI, making it citable.

GitHub has a simple integration with the Zenodo digital repository that allows you to archive a snapshot of code files at a click. You can archive code hosted in the University GitLab or another repository, but the process is a bit more 'manual': you would need to export your code files and then deposit them into any suitable repository along with relevant documentation and a licence statement. Most general data sharing services such as Zenodo and figshare are suitable for archiving code files. You can also deposit code files in the University Archive.

As a means of advertising your code and encouraging re-use you may want to consider publishing a software paper. This will also give you another citable output. There are plenty of journals that will publish software papers.

Appendix. Funders' data management requirements

This list is not exhaustive. Wherever a funder requires a DMP as part of an application, this should be sent to the Research Data Manager for review prior to submission. UoR DMP guides are published at on the Research Data Management Website.

FUNDER	REQUIREMENTS	POLICY URL	UOR GUIDANCE
AHRC	2 page DMP attachment	https://www.ukri.org/publications/ahrc-research-funding-guide/	Yes
BBSRC	1 page DMP attachment	https://www.ukri.org/councils/bbsrc/guidance-for-applicants/what-to-include-in-your-application/data-management-plan/	Yes
British Academy	Deposit of datasets statement	Application form	
Cancer Research UK	Applicants to Population Research Committee: DMP attachment Applicants to other committees: free-text DMP	https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy	
EPSRC	No DMP required but research data policy applies to grant-holders	https://www.epsrc.ac.uk/about/standards/researchdata/expectations/	Yes
ESRC	3 page DMP attachment	https://www.ukri.org/publications/esrc-research-data-policy/	Yes
Horizon Europe	0.5-1 page addressing mandatory open science practices including research data management	https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/9570017e-cd82-11eb-ac72-01aa75ed71a1	Yes

FUNDER	REQUIREMENTS	POLICY URL	UOR GUIDANCE
MRC	3 page DMP attachment	https://www.ukri.org/councils/mrc/guidance-for-applicants/policies-and-guidance-for-researchers/data-sharing/	
NERC	Up to 1 page Outline DMP; Full DMP to be developed with nominated NERC data centre post-award	https://nerc.ukri.org/research/sites/data/dmp/	Yes
Newton Fund	200 words Outline of data management and data sharing plan	See Scheme Notes for guidance	Yes (see Royal Society)
Parkinson's UK	500-word Data Management and Sharing Plan	https://www.parkinsons.org.uk/research/data-management-and-sharing	
Royal Society/ Newton Fund	200 words/1500 characters (depending on scheme) Outline of data management and data sharing plan	See Scheme Notes for guidance	Yes
STFC	2 page DMP attachment	http://www.stfc.ac.uk/funding/research-grants/peer-review-and-assessment/data-management-review-guidance/	
Wellcome Trust	Outputs management plan, including data and software	https://wellcome.org/grant-funding/guidance/how-complete-outputs-management-plan	