

Data Management Planning Guidelines

A data management plan is a formal document developed at the start of a research project which outlines all aspects of your data in terms of;

- what data will be created
- how will this data be generated
- what are the plans for sharing and preservation of the data and;
- the restrictions that may need to be applied.
-

The University encourages all researchers (including postgraduate students) to include in their research proposals a research data management plan (where appropriate to the study) that explicitly addresses data capture, management, integrity, confidentiality, retention, sharing and publication.

The University of Edinburgh offers a free online course which helps to explain the way in which data management should be undertaken and gives guidance on writing plans:

<http://datalib.edina.ac.uk/mantra/>

Data management planning tool

DMPonline (<https://dmponline.dcc.ac.uk/>) is a flexible web-based tool to assist users to create personalised plans according to their context or research funder. You can create plans with DMPonline to meet funder requirements at the grant application stage, or to support data management throughout your research project. DMPonline also contains useful guidance on how to prepare a data management plan and carry it through to execution.

The following guidance has six sections and is based on the [Research Data Management Policy](#). For each question there is an example question and guidance¹.

Data Capture

What data will be generated or reused in this research?

Guidance

Here you should give a short description of the data you will create, or any third-party data that you plan to reuse. Consider the types of data your research will generate, for example qualitative survey data, computational models, statistics, measurements, text, images, audio visual data, or samples. Also consider the longevity of the file formats that you choose. Using open standards so your data can be read by a variety of software makes it easier to preserve and share with others. If you are reusing third-party data, make sure you have sought appropriate permissions to allow for any data sharing or preservation plans.

Example answer

“Every two days, we will subsample E. ojjinis populations growing under our treatment conditions. We will use a microscope to identify the life-stage and sex of the subsampled individuals. We will document the information first in a laboratory notebook and then copy

the data into an Excel spreadsheet. For quality control, values will be entered separately by two different people to ensure accuracy. The Excel spreadsheet will be saved as comma-separated value (.csv) file daily and backed up to a server. After all data are collected, the Excel spreadsheet will be saved as a .csv file and imported into the program R for statistical analysis.”

How much data will be generated?

Guidance

The University has high quality storage with guaranteed backup and resilience through the usual personal drive space (M drive) and on shared drives on the network. If you are unsure of your specific data storage needs then please speak to IS&T as soon as possible for advice. It may also be helpful to speak to the Electronic Resources Librarian for advice on your data set may be best captures in the institutional repository.

Data Management

How will the data be documented to ensure it can be understood?

Guidance

Producing good documentation and metadata provides context for your data, and makes it easier to find and use in the long term. The amount of effort put into documenting your data will depend on the intended lifespan and how broadly you intend to share it. You may want to capture details about what instrumentation has been used and how that has been calibrated, full variable and value labels, and details about your methodology. Some of this information may be captured already in lab notebooks, project documents or research papers.

Example answer

"Metadata will be tagged in XML using the Data Documentation Initiative (DDI) format. The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively". Source: ICPSR.

Where will the data be stored and backed-up?

Guidance

The University recommends that you store your research data on usual network servers that that are maintained by IS&T. This ensures that the data are held securely and are regularly backed-up.

Example answer

“The data will be stored on University of Bolton servers. This is high quality, enterprise-class storage with guaranteed backup and resilience.”

Integrity

How will you quality assure your data?

Guidance

Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.

Example answer

"New data produced during the project will be derived from existing spatial and statistical datasets. These will be highly reliant on the quality assurance procedures of the agencies producing the original data. Information on the quality and reliability of the existing spatial data utilised by the project will be included in the SECRA meta-data. This will also indicate the data source and any procedures (either spatial or statistical) that have been applied to the data layers. If appropriate, statistical data showing the variation (standard deviation) in the data created will be included in the spatial database. Data modelled by the project will be checked for internal consistency by the research team. This will involve sampling data for various points to check that any algorithm applied has produced the expected and desired outputs. The data will then, if appropriate, be checked against a similar or surrogate variable from another source to determine if the results are comparable with existing data from other agencies."

Source: RELU Social and environmental inequalities in rural areas, University of Edinburgh.

Confidentiality

How will you manage any ethical and intellectual property issues?

Guidance

Consider whether you will be creating any sensitive data that need to be kept confidential to protect participants or to safeguard intellectual property. If so, what procedures will you follow, for example obtaining permissions via consent agreements and data licenses, or protecting information via anonymisation, restricted access and embargo period.

Example answer

"Some consultancies or individuals may be reluctant to have their data shared. Such concerns will be resolved for the interviews and focus groups by ensuring all transcripts are anonymous and identifying detail (e.g. campaign names) are removed or given pseudonyms in the transcript. The case analyses, industry awards data and analysis of promotional material will proceed on the basis of publicly available documents, so the issue of sharing should not arise. However, copyright may be an issue for some of these documents. To resolve this, and before data is stored by ESDS, copyright permission will be sought for web and other promotional material written by participating organisations".

Source: Leeds RoadMaP.

Retention and Preservation

Which data do you plan to keep and for how long?

Guidance

Consider which data are of long-term value and should be retained. You should keep the data needed to validate any published findings, and may wish to keep other data that could be reused.

How long do the data need to be kept?

Check if you are subject to legal obligations or expectations from your funder as these may have precedence over any other recommendations. The UKRIO Code of Practice for Research section 3.12 recommends a minimum of 3 years

(<http://www.ukrio.org/publications/code-of-practice-for-research/3-0-standards-for-organisations-and-researchers/3-12-collection-and-retention-of-data/>).

Example answer

"Our intent is that the long-term high quality final data product generated by this project will be available for use by the research and policy communities in perpetuity. The raw supporting data will be available in perpetuity as well, for use by researchers to confirm the quality of the Mauna Loa Record. The investigators have made arrangements for long-term stewardship and curation at the Carbon Dioxide information and Analysis Center (CDIAC), Oak Ridge National Laboratory".

How will the data be preserved?

Guidance

Consider how your data will be managed so it can still be opened and understood in several years' time, and who will do any work needed to ensure this? Will you offer your data for deposit in a subject-specific data archive in addition to the University's own institutional repository (<http://ubir.bolton.ac.uk/>)? The data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the research grant.

Example answer

"The University of Bolton undertakes to maintain the digital outputs of this project for the long-term and will utilise University infrastructure (namely the University of Bolton Institutional Repository [UBIR]) to ensure preservation and continued access. The UBIR is an online repository of multi-disciplinary research datasets hosted by the Library. It acts as a trusted repository, ensuring that research data will be preserved."

Sharing and Publication

Which data will be shared and how?

Guidance

The University expects data to be made available for access and re-use where possible, and provides a data sharing guide to advise on different options. The University's repository (<http://ubir.bolton.ac.uk/>) can be used to share data, as can services from your funder, publisher or research community.

Example answer

"The final data product will be released to the public as soon as the recalibration of standard gasses has been completed and the data have been prepared, typically within six months of collection. There is no period of exclusive use by the data collectors. Users can access documentation and final monthly CO2 data files via the Scripps CO2 Program website(<http://scrippsco2.ucsd.edu>). The data will be made available via ftp download from the Scripps Institution of Oceanography Computer Center. Raw data (continuous concentration measurements, weather data, etc.) will be maintained on an internally accessible server and made available on request at no charge to the user".

Are any restrictions on data sharing required?

Guidance

Not all data can be shared. You may not be able to share your data if:

- it has financial value or is the basis for potentially valuable patents,
- it contains sensitive, personal information about human subjects,
- parts of the data are owned by others.

Outline any restrictions you need to place on data sharing, for example, if the data can only be shared under certain conditions or an embargo period is needed.

Example answer

"Even though the final dataset will be stripped of identifiers prior to release for sharing, we believe that there remains the possibility of deductive disclosure of subjects with unusual characteristics. Thus, we will make the data and associated documentation available to users only under a data-sharing agreement that provides for (1) a commitment to using the data only for research purposes and not to identify any individual participant; (2) a commitment to securing the data using appropriate computer technology; and (3) a commitment to destroying or returning the data after analyses are completed."

Source: NIH data sharing plan examples:

http://www.grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

DATA MANAGEMENT PLANNING GUIDELINES	
Policy ref:	
Version number	1.0
Version date	March 2015
Name of Developer/Reviewer	Andrew Graham
Policy Owner (Group/Centre/Unit)	Research & Graduate School
Person responsible for implementation (postholder)	Designated officer in the R&GS
Approving committee/board	Board of Studies for Research Degrees
Date approved	
Effective from	May 2015
Dissemination method e.g. website	Website
Review frequency	3 years
Reviewing committee	Board of Studies for Research Degrees
Consultation history (individuals/group consulted and dates)	As above
Document history (e.g. rationale for and dates of previous amendments)	

ⁱ The model at the University of Edinburgh has been used to guide the writing of this document and examples taken from: http://www.ed.ac.uk/polopoly_fs/1.160897!/fileManager/Edinburgh_DMP_template_web.pdf