Data Management Planning

# NERC funding applicants

Version 2.6 June 2021

University of Bristol

## Research Data Service

Image: ammonit-591874 1280.png, Pixabay, Public Domain

University of BRISTOL

# SUMMARY

- A one-page (or less) Outline Data Management Plan (ODMP) is required at the application stage.[1]
- A fuller Data Management Plan (DMP) must be provided to NERC within three months of the project's starting date.
- NERC provides a Data Value Checklist[2] to help researchers decide which datasets have long-term value.
- At the end of a research project NERC requires that all datasets with long-term value should be made available for others to use with as few restrictions as possible, and in a timely manner, usually via one of the NERC Data Centres.
- Researchers are entitled to 'right of first use' (i.e. exclusive access) to the data they generate, but this period must not be longer than two years from the end of data collection/creation.
- All research publications arising from NERC funding must include a statement on how underpinning research datasets can be accessed.

# INTRODUCTION

The Natural Environment Research Council (NERC) Data Policy, as part of UK Research & Innovation (UKRI, previously RCUK), emphasises the need for openness and access to the data that underpin research publications. Research data produced by activities funded by the NERC is considered to be a public good which should be made openly available for anyone to use. The same policy includes a formal requirement for all funding applicants to submit a very short Outline Data Management Plan (ODMP) and, if successful, a fuller Data Management Plan; in partnership with one of the NERC Data Centres (see Appendix 1 – NERC Data Centres, below).

The NERC Data Policy[3] applies to all environmental data acquired, assembled or created through activities that are either fully or partially funded by NERC. The Policy also applies to environmental data managed by NERC, but for which NERC was not the original funder. NERC defines environmental data as items or records that are usually obtained by measurement, observation or modelling of the natural world and the impact of humans upon it. This includes data generated through complex systems, such as information retrieval algorithms, data assimilation techniques and the application of numerical models. Separate guidance is available from each Data Centre covering preservation of model code and output[4].

NERC is committed to safeguarding the availability of research data which has long-term value for research, teaching and wider uses, in order to:

- support the integrity, transparency and openness of research;

[1] NERC ODMP Guidance Notes, https://nerc.ukri.org/research/sites/data/dmp/dmp-guidance/
[2] NERC Data Value Checklist, nerc.ukri.org/research/sites/data/policy/data-value-checklist/
[3] NERC Data Policy nerc.ukri.org/research/sites/data/policy/data-policy/
[4] NERC supported and contracted Data Centres, https://nerc.ukri.org/research/sites/data/

- assist in the formal publication of datasets and enable the tracking of their usage through citation and data licences;
- abide by relevant legislation and government guidance on the management and distribution of environmental information;
- ensure the long-term availability of environmental data by supporting several Data Centres (see Appendix 1 – NERC Data Centres) and by stipulating several conditions relating to data sharing, which all recipients of NERC funding must observe.

The NERC stance on the management and sharing of research data is shared by most major research funders, the National Science Foundation and the European Commission.

For more general information concerning research data management issues, please refer to our Brief Guide to Managing Research Data.[5]

## Researcher responsibilities

At the end of a research project NERC requires that all datasets with long-term value should be made available for others to use with as few restrictions as possible, and in a timely manner.

Researchers are entitled to 'right of first use' (i.e. exclusive access) to the data they generate, but this period must not be longer than two years from the end of data collection/creation.

All research publications arising from NERC funding must include a statement on how underpinning research datasets can be accessed. Such supporting research data will usually be made available through one of the NERC Data Centres.

These stipulations apply to all applications for funding, including fellowships and research activities only part-funded by NERC. Researchers funded by NERC who do not meet these requirements may have award payments withheld or become ineligible for future NERC funding.

## Models

NERC recognises that model code and the resulting model data are valuable research outputs, and should be preserved along the same lines as other types of research data. Model code for NERC-funded research should meet the following minimum requirements:

- developed in an open-source environment, where possible;
- governed by a development tool with version control such as subversion or GIT;
- available in a non-proprietary format for storage;
- adequately documented.

Minimum requirements for model input or configuration files are as follows:

- preserved in standard formats (e.g. netCDF);
- governed by a development tool with version control such as subversion or GIT;
- adequately documented.

[5] A brief guide to research data, http://bit.ly/2okoR7W

In order to be 'adequately documented', model documentation should follow the NERC metadata standards for models[6]. At a minimum, documentation should include details of the model, input data, any pre- or post-processing software that was used along with version information, the date when the model output data was created, and the people and institutes responsible for running the model. Model code, and input or output data should be provided to the appropriate NERC Data Centre for preservation at the end of a project.

## Outline Data Management Plan (ODMP)

NERC provide a template ODMP[7]. You are required to state whether or not you intend to create any data, which of the NERC Data Centres you intend to use, and to provide a brief list of any datasets you know you will create.

## Data Management Plan (DMP)

Once you have successfully acquired research funding, your ODMP will be used (in conjunction with the most appropriate NERC data centre) to help produce a fuller and more detailed DMP. The main purpose of the full DMP is to ensure that datasets of long-term value are deposited with the Data Centre in an appropriate format and along with the necessary metadata. The full DMP must be produced within three months of the project's starting date.

Your full DMP will expand on the following areas: backup and security, metadata and documentation, data management responsibilities (for example, who is responsible for capturing data in the field, producing metadata, transferring metadata and data, and how version control will be achieved) expected sizes and formats of datasets, potential challenges relating to data transfer or re-usability (such as exceptional size or complexity), plans for data preservation, and details of any existing datasets to be used during the project.

## Metadata

Metadata is 'data about data' and is information (or cataloguing information) that enables data users to find and/or use a dataset. In your DMP you should outline plans for documenting your research data, to meet both your own needs and those of later users.

In attempting to organise and document your data it may help to imagine a secondary data user trying to make sense of your data in your absence, after the end of your project. If presented with only the data itself, a secondary user may be faced with the difficult task of 'unpicking' it. How will they make sense of your file and folder naming conventions? Has any special software been used to create your data? What extra information would they need to make maximum use of it?

For more information on relevant metadata standards, including the NERC metadata standards for models, contact the relevant NERC Data Centre for your subject area.

[6] NERC metadata standards for models, http://model-search.nerc.ac.uk/about

[7] NERC Data management planning, nerc.ukri.org/research/sites/data/dmp/

## Data storage

It is recommended that, as you create data, you store it in the University's own Research Data Storage Facility (RDSF), managed by the Advanced Computing Research Centre (ACRC).[8] Each research staff member is entitled to 5TB of storage without charge. If your storage quota is used up, or your project requires more storage space, there will be a cost and ACRC should be contacted for guidance before your application is finalised. The back-up procedures, policies and controlled access arrangements used by the RDSF are of a very high standard. If you do not intend to make use of the RDSF, your storage provider's back up procedures should be briefly described instead.

Your DMP should briefly indicate how you'll keep your data safe before it's deposited in a storage facility such as the RDSF. This is particularly important if you're conducting field research. As a minimum requirement, try to ensure that at least two copies of the data always exist, and that every copy can easily be accounted for and located if required.

If you expect to need any specialised help with creating or managing your data, such as help with database design, you should also mention this in the DMP.

## Data quality

Your DMP should describe how you will ensure the quality of your research data. Quality should be considered whenever data is created or altered, for instance at the time of data collection, data entry or digitisation. It may be appropriate to nominate a research data manager within the team and outline the procedures they will use to ensure data quality, such as dedicated time to check data, entering values into pre-prepared databases, or using templates.

If you plan to integrate student data into your datasets, you should mention this within the DMP.

## Ethics, IPR and data protection issues

NERC expects funding applicants to investigate any likely ethical or Intellectual Property Rights (IPR) issues that are likely to affect your ability to share your data, and these should be mentioned in the DMP. If you are planning to use existing data as part of your research, the data may be subject to certain copyright or other restrictions that could prevent you from sharing any new data you derive from them. You should give full and appropriate acknowledgement, via citation, for any existing data you expect to use.

Unless stated otherwise, the ownership of intellectual property lies with the organisation carrying out the research. However, if you plan to work collaboratively with an external partner, copyright and IPR issues may need to be clarified in a formal agreement. While this isn't required as part of your application, it should be mentioned that, if the application is successful, such an agreement will be created. The University's Research Enterprise and Development (RED)[9] can advise further on collaborative research agreements and other IPR issues.

[8] Advanced Computing Research Centre, http://www.acrc.bris.ac.uk

[9] Research Enterprise and Development, http://www.bristol.ac.uk/red/contracts

All recipients of research grants must abide by the Data Protection Act 2018 and the General Data Protection Regulation (GDPR). If you plan to handle sensitive and/or personal data, extra security measures must be considered. The Office of the University Secretary[10] can provide advice on observing data protection legislation.

## Table of datasets

It may be difficult for you to predict accurately the nature and extent of the datasets your project will generate, therefore NERC only requires you to make an estimate at the funding application stage. You won't necessarily need to mention everything, only the most significant datasets that are likely to have long-term value. If you are uncertain whether a dataset is likely to have long-term value or not, it may help to look at the NERC Data Value Checklist[11] (see below). Although this tool is primarily intended to be used when preparing a more detailed data management plan (more about this below) you may also find it useful during the process of creating a DMP.

For each dataset which you intend to generate and which you believe may have long-term value, you should provide the following information in a table:

- Data Centre - the name of the most appropriate NERC Data Centre. If you're unsure which Data Centre is the most appropriate for deposition of your data, visit the Data Centre's own website and read its collections policy. If you're still in doubt, it might help to send the Data Centre/s concerned a brief description of your dataset and ask their opinion on its suitability for deposition. Individual projects can contribute to more than one Data Centre. (See Appendix 1 for a list of NERC Data Centres)
- Dataset description - a brief (one or two sentence) description of the data. Examples might be 'photographs of field area' or 'raw broadband magnetotelluric data'.
- Release date for giving data to Data Centre - if you don't have a specific date, you can specify a period such as 'by the end of the project' or 'during year two'. It is expected that data should be delivered to a NERC Data Centre within two years of end of data collection.
- Re-use scenarios - if you have an idea of the type of secondary user who might make use of your dataset, describe them here in one or two sentences. Examples might be 'oceanographic researchers' or 'commercial researchers'.

## Assessing data value

The NERC Data Value Checklist is a tool to help you assess the long-term value of a dataset when preparing a full data management plan.

The Checklist informs all decisions that NERC Data Centres make on the acquisition, preservation and eventual disposal of environmental data. The criteria described in the Checklist do not directly indicate whether the data should be considered 'valuable', but instead offer guidance on assessing long-term value.

[10] Office of the University Secretary, www.bristol.ac.uk/secretary/data-protection

[11] NERC Data Value Checklist, nerc.ukri.org/research/sites/data/policy/data-value-checklist/

*Mandatory criteria* (criteria which require the retention of data) are:

- legal or legislative reasons for data retention (for example, compliance with the Environmental Information Regulations or contractual obligations);
- data that are likely to be the subject of legal challenge or of litigation.

*Important criteria* (criteria which strongly suggest the retention of data) are:

- data that are new and unique;
- data that are irreplaceable (for example, data that arise from observations and sampling rather than repeatable simulations or experiments);
- data that have a broad extent and so are widely re-usable;
- data that are of special scientific or communal importance;
- data which set an important precedent;
- data that are part of a wider, current trend in science;
- data that are likely to meet future needs;
- data which add value to an existing dataset;
- data that have clear potential for reuse;
- data that are likely to be cited within a publication.

*Supporting criteria* (criteria which suggest the retention of data) are:

- data that are 'raw' and unprocessed;
- data that would be expensive to reproduce;
- if the deposited version is likely to be the reference version of the dataset;
- accurate and detailed metadata accompany the data, to support any future re-use;
- more high value data than low value data in the dataset;
- data in a format which supports deposit in a data centre and subsequent storage and preservation;
- permissions are in place to permit data re-use;
- no special software is required to use the data so the data could easily be converted into a more widely used format.

## Data submission and access

The appropriate NERC Data Centre should be provided with a copy of your finalised data as soon as possible after the end of data collection. This will allow the data centre to check that all the necessary information for readily allowing others to re-use the data is included in the documentation. NERC will, however, allow funded researchers a reasonable amount of time to finalise their datasets and publish their findings, during what is known as an 'embargo period'. NERC considers that in most cases a reasonable embargo period is a maximum of two years from the end of data collection. Data submitted to a data centre during an agreed embargo period will remain restricted for the period defined, though many researchers choose not to apply an embargo period and are happy for their data to be made available to others once they have been finalised.

Once your data has been deposited with a NERC Data Centre and made accessible, it will be accompanied by a data licence. In general, all data made available by the NERC Data Centres can be accessed by anyone. However, in the case of some third-party datasets,

there may be restrictions on who can access the data or what can be done with them, and any such restrictions will be made clear when the data are requested. The data licence will also specify that users of the data must acknowledge the originator of the data in any publication or other derived work.

In order to cite datasets that underpin research publications (see Researcher responsibilities, listed above) data may be assigned a Digital Object Identifier (DOI) by a Data Centre. A DOI is a unique identifier that does not change over time and will serve as the 'permanent online address' of a specific dataset. A DOI will also help to support the tracking of data usage through the publication and citation of data sets. In order for the receiving Data Centre to issue a DOI, data must be deposited in good condition, with appropriate metadata and of a suitable level of technical quality. The submitter is responsible for ensuring data meets the required level of quality.

Metadata pertaining to all datasets held within the Data Centres will be made available through the NERC Data Catalogue Service.[12] This service provides an integrated, searchable catalogue of the data holdings of NERC's Data Centres, and can be used to find information on what data the NERC data centres hold and how to access these data.

# CITING RESEARCH DATA IN RESEARCH OUTPUTS

From 1st April 2013 all the UK's research funding councils, as part of UKRI (formerly RCUK), require research outputs (i.e. journal articles) to provide a means by which third parties can access any underpinning research datasets. This may be a reference (such as a unique URL or DOI) printed in a paper which will lead an enquirer to a specific web page where the data is available. Or the enquirer might be directed to a page which displays the contact details of a custodian of the data and asked to email them in order to gain access to the data.

Given the extended timescales involved in publication, it is strongly recommended that the authors of published academic outputs do not provide their current contact details as a means of accessing underpinning research data, as these will change over time. The NERC Data Centres can provide unique reference identifiers for deposited datasets which can be included in publications instead.

[12] NERC Data Catalogue Service, https://data-search.nerc.ac.uk/geonetwork/srv/eng/catalog.search#/home

# APPENDIX 1: NERC DATA CENTRES

| Data Centre | Area of Interest | Website | Contact |
|---|---|---|---|
| British Oceanographic Data Centre (BODC) | Marine science | www.bodc.ac.uk | enquiries@bodc.ac.uk |
| British Atmospheric Data Centre (BADC) – part of the Centre for Environmental Data Analysis (CEDA) | Atmospheric science | archive.ceda.ac.uk | support@ceda.ac.uk |
| NERC Earth Observation Data Centre (NEODC) – part of the Centre for Environmental Data Analysis (CEDA) | Earth observation | archive.ceda.ac.uk | support@ceda.ac.uk |
| UK Solar System Data Centre (UKSSDC) – part of the Centre for Environmental Data Analysis (CEDA) | Solar and space physics | www.ukssdc.ac.uk | support@ukssdc.ac.uk |
| Environmental Information Data Centre (EIDC) | Terrestrial & freshwater science, hydrology and bioinformatics | www.ceh.ac.uk/data | gateway@ceh.ac.uk |
| National Geoscience Data Centre (NGDC) | Earth sciences | www.bgs.ac.uk/services/NGDC | enquiries@bgs.ac.uk |
| Polar Data Centre (PDC) | Polar and cryosphere | www.bas.ac.uk/data/uk-pdc | polardatacentre@bas.ac.uk |
| Archaeology Data Service (ADS) | Science-based archaeology | archaeologydataservice.ac.uk | mhelp@archaeologydataservice.ac.uk |