NERC Data Policy - Guidance Notes

Author: Mark Thorley – NERC Data Management Coordinator

Contents

Guidance notes for the NERC Data Policy

This document provides guidance in applying the NERC data policy, which was formally approved by the NERC Executive Board in September 2010 and is being progressively implemented from January 2011 onwards. These guidance notes will be developed as elements of the policy are fully implemented, and will also evolve to reflect changes in practice and scientific priorities.

1. Data covered by the NERC Data Policy

The NERC Data Policy applies to environmental data acquired, assembled or created through research, survey and monitoring activities that are either fully or partially funded by NERC. It also applies to environmental data managed by NERC, where NERC was not the original funder.

The policy does not cover NERC's value-added information products. In general, all information products held by NERC will be made available for free for non-commercial teaching and research purposes. Full details are in NERC's Licensing and Charging Policy.

2. Definition of terms

a. Environmental data

NERC defines environmental data as individual items or records (both digital and analogue) usually obtained by measurement, observation or modelling of the natural world and the impact of humans upon it, including all necessary calibration and quality control. This includes data generated through complex systems, such as information retrieval algorithms, data assimilation techniques and the application of numerical models. However, it does not include the models themselves.

A few examples of what we mean by environmental data are given below:
- Model output from running a numerical climate model
- Time series logged by environmental instrumentation
- Conductivity-Temperature-Depth casts from oceanographic cruises
- Groundwater chemistry and stable isotope measurements
- Butterfly abundance observations

b. Information products

Information Products are created by adding a level of intellectual input that refines or adds value to data through interpretation and/or combination with other data. They result from analysis or repackaging of data in such a way that has provided significant added value (intellectual or commercial), e.g. tidal predictions or Land Cover maps are not covered by the data policy.

c. Third-parties and third-party data

Third-parties are organisations or individuals that are not part of NERC. Third-party data are data generated by third-parties which have not been funded by NERC. The NERC Data Centres hold third-

party data, which we regard as having long-term value to the environmental science community. Data Centres also hold composite data sets which may include both NERC and third-party data. Sometimes there may be constraints on how these third-party data can be accessed and used, which may also apply to composite data sets.

d. NERC-funded

NERC-funded is defined as where NERC has been a full or a partial funder of the activities that create environmental data. Normally, where NERC is a partial funder of an activity, it will be as part of a joint-funding programme with other organisations. In this case, the funding partners will agree what specific data management conditions will apply, for example, the length of the 'embargo' period. However, all specific conditions will still meet the requirements of NERC's policy.

e. Data Creators

Data Creators are individuals, teams or organisations that collect or generate data as part of NERC-funded activities, including scientists in Higher Education Institutes and NERC's Research Centres.

f. Data Users

Data Users are all those using data supplied by the NERC Environmental Data Centres.

g. Environmental Data Centres

These are facilities designated by NERC to manage and disseminate the environmental data that it holds.

3. Access to data – Information for Data Users

This section provides information to those who wish to use environmental data held by the NERC Data Centres and provides guidance for Data Policy points 1-6.

a. Restrictions to access

Anyone is allowed to access NERC-funded data, regardless of the purpose for which they intend to use them, including commercial gain.

In general, all data made available by the NERC Data Centres can be accessed by anyone for any purpose. However, for some third party data sets there may be restrictions on who can access or what can be done with the data, and any such restrictions will be made clear when the data are requested.

i. Freedom of Information and Environmental Information Regulations

NERC is a Public Authority and, as such, is subject to freedom of information legislation. The environmental data held by NERC's data centres is classed as 'Environmental Information' under the terms of the Environmental Information Regulations 2004 (EIRs), and there is a presumption that these data will be available to all who request them unless there is a clear reason, supported by the EIR exceptions, as to why they should be withheld.

NERC's Environmental Data Centres will try to make the data they hold as open and accessible as possible in line with the requirements of the EIRs. However, there are times when the data centres

will have to restrict access to data. Such restrictions are normally used to protect the research process allowing researchers a reasonable time to work up their data sets and publish their findings, or, when a release of data might lead to damage to the environment. Please talk to the data centres if you have concerns over access, as they will try to be as flexible as possible.

All restrictions on access that NERC uses are fully supported by the Exceptions to Access defined by the Environmental Information Regulations 2004. The EIRs also provide a process of appeal if you consider that NERC is unreasonably withholding data. Further information on this appeal mechanism can be found on the freedom of information pages of the NERC web site (foi.nerc.ac.uk).

ii. Embargo periods
A key reason to restrict access is to protect the research process by allowing researchers a reasonable amount of time to work-up their data sets and publish their findings. This 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. What this means in practice is that you may have to wait until an embargo period has expired before you can access the data, though it may be possible to negotiate an earlier release.

iii. Long-term time series
Long-term time series data are a special case – whilst early data may be publicly accessible, more recent data may be under an embargo period. This is to allow the recent data to be worked-up and for findings to be published. Data centres will provide specific details on any restrictions for each time series.

b. Data licences
Data supplied by the NERC data centres will be accompanied by a data licence. A data licence is an agreement between NERC and a data user; it outlines any limitations on how the data may be used, how the source/creator of the data must be acknowledged and the limits of NERC's liability for the data it provides. NERC data licences are based on the UK Open Government Licence for Public Sector Information.

The data licence will request that you do not pass data supplied by NERC on to third parties, but to refer the third party to the data centre. This is to ensure the third party receives the most recent version of the data, and also allows the data centres to provide the data originator with information on how their data are being re-used.

c. Acknowledgement
Those who use data supplied by NERC must acknowledge in any publication or any other derived work, the contribution made by those who have created and worked up the data. The NERC data licence will contain information on how best to do this.

d. Charges for data
In the majority of cases, environmental data provided by the NERC data centres will be supplied for free, however, there are some exceptions.

To support provision of equal access to data to all users, where a request is large or complex the data centres will make an appropriate administrative charge on a cost recovery basis. A large or

complex request is where a significant amount of manual intervention is required from data centre staff, or data are requested on non-standard media or in a non-standard format.

Data centres have to charge for some third-party data sets that they hold, or composite data sets that contain third-party data, where the third-party's conditions of access stipulate this.

The data centres will supply information on any charges that may apply before a request is met.

e. Where to find data

i. NERC Data Catalogue Service
The NERC Data Catalogue Service (data-search.nerc.ac.uk) provides an integrated, searchable catalogue of the data holdings of NERC's data centres. It can be used to find information on what data the NERC data centres hold and how to access these data.

ii. NERC Environmental Data Centres
The NERC Environmental Data Centres have searchable catalogues and data download facilities on their web sites that have been designed to serve their individual science communities. The data centres also have enquiries desks to help with more bespoke data requirements.

More information about the NERC Environmental Data Centres is given in Section 5 and a complete list of the data centres and links to their web pages can be found at data.nerc.ac.uk.

4. Generating new NERC-funded data – Information for Data Creators
This section provides information to those collecting data as part of NERC-funded activities, and provides guidance for Data Policy points 11-15.

Good data management techniques are a fundamental component of good scientific practice. Data management planning provides a mechanism to define key data management activities which are necessary to ensure the integrity and security of the data sets generated by the research process.

The policy requirements for identifying data of long-term value and for developing data management plans have yet to be formally implemented. This is to allow time for NERC to develop the necessary process and mechanisms for reviewing and managing data management plans, and to work with the NERC community to develop criteria to help identify data of long-term value. This area of the policy will be implemented over the next 12 to 18 months. Only when fully operational will NERC consider the application of sanctions for non-compliance.

During this implementation period, NERC grant holders and staff within NERC's research centres are required to offer to deposit with a NERC data centre copies of any data sets resulting from the research supported. The data centres will take copies of those data sets considered to be of long-term value, which will be identified in discussion between the researcher and the data centre. Appropriate data formats and metadata requirements will be agreed with the data centres as part of this process.

a. Identifying the relevant NERC data centre

It should usually be obvious from the discipline area you are working in which is the NERC data centre relevant to your area. If you are uncertain or if your research area is multi-disciplinary, then we will need to work with you to identify the appropriate data centre. Contact data@nerc.ac.uk.

For Research Programmes, a lead data centre will be identified, which will be the primary point of contact for data management issues within the programme. There may be several data centres that receive data from projects within a programme, but this activity will be coordinated through the lead data centre.

Further information on the data centres and their discipline areas is available in Section 5 and at data.nerc.ac.uk.

b. Timescales for data submission

The timescale for submission of data sets should be agreed in advance with the data centre at the outset of the project.

In general, data centres should have a copy of the finalised data as soon after the end of data collection as is possible. The end of data collection is defined as the point at which the data become available from an instrument or experiment. It does not mean the end of the NERC-funded project.

Submission of data as soon as a finalised version is ready allows the data centre to ensure that all the necessary information to allow others to readily re-use the data are documented whilst the detailed knowledge of the specifics of the data set are still fresh in a researcher's mind. It is neither in the researcher's nor the data centre's interests to have to remember and document detailed specifics of data sets several years after the data were collected.

c. Embargo periods

NERC will allow researchers a reasonable amount of time to work-up their data sets and publish their findings. This is known as an 'embargo period', and NERC considers that, in most cases, a reasonable embargo period is a maximum of two years from the end of data collection. However, in exceptional circumstances, a longer-period may be applicable, for example if it can be justified by the requirements of the research. The specific embargo period must be agreed in advance with the data centre.

Once the embargo period has expired, the data will be made available to anyone who requests them, unless there is a valid reason to restrict access under the EIRs.

Data submitted to a data centre whilst under an agreed embargo period will remain restricted for the defined period. It should be noted that 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 worked up.

An embargo period starts from the end of data collection, e.g. the end of an oceanographic cruise, the point at which the data are downloaded from the instrument, or the point at which the data has been gathered from an experiment or analysis. It is not from the point at which the data were submitted to the data centre or the end of the NERC-funded project.

For specific research programmes there may be a requirement for a reduced embargo period between projects within the programme, to enable inter-project working for delivery of the overall programme objectives. However, wider sharing of the data will still be restricted until the overall programme embargo period has expired.

Once an embargo period has expired the data are available to anybody to use for whatever purpose. However, there may be cases where a data set generated through NERC funding contains data supplied by a third-party which may have restrictions that apply to their use. The NERC data centres will provide advice on third-party data issues on a case-by-case basis.

d. Data formats

It is not possible for these guidance notes to cover all the potential formats for all possible data types. Therefore, this is something that needs to be discussed and agreed with the data centre at the outset of the project.

e. Metadata

Metadata are a crucial part of any data archive since they ensure that the data can be understood at a later date. It is essential that metadata are submitted at the same time as the data sets to which they pertain. The data centres require detailed information about how the data were arrived at, i.e. metadata, covering methods of collection, processing, calibration and quality control must be supplied, so that all the necessary information is available to allow others to effectively re-use the data.

The term metadata encompasses all the information necessary to interpret, understand and use a given data set. Discovery metadata particularly apply to information (keywords) that can be used to identify and locate data sets that meet the user's requirements via a Web browser or a Web based catalogue, for example. Detailed usage metadata include additional information necessary for a user to work with the data without needing to refer back to the data provider. The metadata required by data centres include both discovery and detailed metadata.

Metadata pertaining to observational data, for example, include details about **how** (with which instrument or technique), **when** and **where** the data have been collected, by **whom** (including affiliation and contact address or telephone number) and in the framework of which research project.

In the case of all submitted data, the data centre needs to know how the values were arrived at. The derivation process must be stated: all processing and calibration steps should be described and calibration values supplied. The nature and units of the recorded variables are essential, as well as the grid or the reference system. As much information as possible about fieldwork instrumentation should be included, e.g. serial number, copies of manufacturer's calibration sheets, and recent calibrations.

Metadata pertaining to model output should include the name of the model, the conditions of the calculation, the nature of its output, the geographical domain over which the output is defined (when applicable). Specific conditions applying to the model or the experiment may be mentioned. Metadata also include information on the format in which the data are stored, and the order of the variables, to allow potential users to read them. Metadata pertaining to software models include the key points of the theory on which the model is based, the techniques and computational language used, and references.

Detailed subject specific guidance on all aspects of metadata creation is available from the data centres.

f. Intellectual property rights

Intellectual property rights (IPR) in the data that a researcher generates depends on who a researcher works for and their contract of employment. It is normally the employer of the researcher that owns the IPR. If you work for NERC, the IPR belong to NERC. If you work for a university, the majority of the time the IPR will belong to the university, but, this does depend on your contract of employment.

The requirement to deposit data with a NERC data centre does not affect intellectual property rights. Where NERC does not own the IPR, the rights owner will be required to grant to NERC a non-exclusive licence to allow NERC to manage and supply the data for reuse. NERC is developing a 'Data Deposit Licence' which will form the basis of this rights assignment process.

g. Access by others to data generated by NERC funding

NERC and the majority of organisations that it funds are subject to the UK's Freedom of Information legislation, specifically the EIRs. This legislation provides a set of rules which NERC and other public bodies, such as universities, must follow in deciding if data or information can be made publicly available. Regardless of where a researcher is based, the chances are that they will be working for a public body, and thus anyone can request the data they hold using Freedom of Information legislation.

The EIRs do not require that all environmental data have to be made available in all cases. However, there is a presumption that it is in the public interest to release data unless there is a good reason otherwise. This means that a NERC-funded researcher cannot arbitrarily decide who can and cannot have copies of the data they have collected.

For the non-specialist dealing with requests for data under FoI and EIR legislation can be complex and daunting. The NERC data centres have expertise in this area and will handle any requests for data they hold. This removes the burden from researchers and helps them to meet their obligations under the legislation. This means that if a request comes to a researcher for data they have deposited in a NERC data centre, they are free to refer the request on to the data centre who will deal with it on their behalf.

h. Acknowledgement

All data supplied by NERC will be accompanied by a data licence. This will specify that users of the data must acknowledge the originator of the data in any publication or other derived work. However, NERC cannot guarantee that users of the data will do this.

In the longer-term, NERC and others working in research data management are looking to develop mechanisms to support the tracking of data usage through the publication and citation of data sets.

i. Sanctions

The requirements for data management activities described in the NERC Data Policy, and detailed in these Guidance Notes, form part of the terms and conditions applying to NERC award holders and their research institutions.

It is NERC's intention that once the Data Policy is fully implemented, those who do not meet the data management requirements described within points 11-13 of the policy risk having award payments withheld or becoming ineligible for future funding.

5. NERC Data Centres

Environmental data can be difficult and expensive to collect and often irreplaceable; therefore it makes sense to ensure that they are preserved and useable over the long-term. This enables the maximum value to be gained from the investment of public funds made in their collection.

The network of NERC data centres are responsible for the long-term curation of data and provide access to NERC's data holdings. By submitting data to the data centre, data providers will meet their obligations under NERC's Data Policy, enabling NERC to provide the widest possible opportunities for re-use of these data. Data centres also provide support and guidance in data management to those funded by NERC. Specific activities of the NERC Environmental Data Centres include:

- Maintaining a comprehensive and integrated catalogue of their environmental data holdings. Having data advertised within this catalogue helps promotes re-use, thus increasing the opportunities for acknowledgement and citation of data;
- Ensuring that high quality metadata are stored with the data so that users have all the information they need to use the data in the future without normally having to refer back to the data creator;
- Handling any requests for data, removing the burden from data providers and helping them to meet their obligations under FoI legislation. This means that if a request for data comes directly to them, they are free to refer the request on to the data centre for processing;
- Providing a secure back-up for data, should a data provider's copy become lost or corrupt.

The data centres and their scientific disciplines are:

- British Atmospheric Data Centre (BADC) - Atmospheric science
- National Geoscience Data Centre (NGDC) - Earth sciences
- NERC Earth Observation Data Centre (NEODC) - Earth observation
- British Oceanographic Data Centre (BODC) - Marine Science
- Polar Data Centre (PDC) - Polar Science
- Environmental Information Data Centre (EIDC) - Terrestrial & freshwater science, Hydrology and Bioinformatics
- UK Solar System Data Centre – Solar-terrestrial physics
- Archaeology Data Service (ADS)* - Science-based archaeology

*Note, ADS is not a NERC data centre, but it is supported by NERC for the long-term management of NERC-funded science-based archaeology (SBA) data, and as such, the same requirements apply to SBA data as all other NERC-funded environmental data.

For general enquiries contact the NERC data management co-ordinator at data@nerc.ac.uk.

a. Charges for deposit of data

NERC supports its data centres to provide for the long-term, post-project management of data generated through NERC-funded activities. Currently NERC-funded researchers are required to include in grant applications the costs for their project specific data management activities, including the data centres costs related to taking in data of long-term value.

The data centres are also happy to take relevant environmental data of long-term value from third-parties; however, they may have to charge for the costs of ingestion and/or storage of these data, and any charges are by negotiation with the relevant data centre.

6. Access to data underpinning research publications

The following provides guidance on Data Policy points 16 & 17.

Data that underpin research publications are the data that have been analysed and reported on within a publication. The NERC Data Policy requires all research publications that arise from NERC funding to include a statement on how the supporting data and any other relevant research materials can be accessed. By having a statement on data and research materials, NERC is looking to ensure that the research it funds is transparent and reproducible, to allow others to confirm or challenge the research.

An example statement:

"The following data sets have been used in the production of this paper…. These data are available from the British Atmospheric Data Centre."

The data policy also requires NERC's own staff to make the supporting data for all their publications available through the NERC Data Centres. This policy requirement will be implemented by the research centres during the next 12 to 18 months.

The underpinning data and research materials should be preserved and accessible for a minimum of 10 years after completion of the research. However, for projects of major importance, this may need to be 20 years or longer. See the RCUK Policy and Code of Conduct on the Governance of Good Research Conduct: Integrity, Clarity and Good Management. July 2009. See www.rcuk.ac.uk/Publications/researchers/Pages/grc.aspx