



**NATURAL
ENVIRONMENT
RESEARCH COUNCIL**



MACRONUTRIENT CYCLES (MC) PROGRAMME DATA POLICY

Summary

This policy aims to facilitate collaboration within and across the various projects, to protect the researchers' rights to publish their results within a reasonable time period (the right of 'first use'), and to further the dissemination, exploitation and scientific impact of results obtained by the programme's research projects.

Proposals will be expected to outline plans for the management of data created, acquired, stored and used by the project. PIs should identify the key data outputs, of long-term value to the community, to be submitted for long-term storage by an appropriate NERC Environmental Data Centre. As there is to be no top-slicing of the MC programme, these costs should include, as two separate components, a) project participants' costs of managing data within the project itself with preparation of data for storage; b) costs of the Data Centre(s), incurred for liaising with the project, assessing long-term data storage requirements and providing support on data management, quality assurance, and the conditioning (formatting etc.) of data to be retained. Costs should be based on Data Centre's effort required and daily payment rate. Applicants should contact the relevant Data Centre to establish likely costs. Data Centre costs will be a 'notional' part of the award, and will be provided directly to the Data Centres, rather than via the project, similar to Services and Facilities costs.

1 Aims and scope

The MC Programme aims to understand the processes controlling nitrogen, carbon and phosphorus in the environment and to develop experiments and models to assist the scientific understanding, the interactions occurring under a changed climate, the impacts of environmental change and to provide predictive methodologies for stakeholders. Achieving this aim will mean using available observations more comprehensively as well as collecting additional data to support process science and modelling. The effective management of data generated by the MC programme supported activities is an important priority.

The purpose of the MC programme Data Policy is to provide rules and guidelines to help the researchers supported to manage their data effectively in order to help meet the scientific aims and NERC's Data Policy obligations. Appendix 1 gives a summary of the Data Policy which was approved by NERC Council for implementation in January 2011

www.nerc.ac.uk/research/sites/data/policy/ The MC programme Data Policy document does not replace the NERC Data Policy: it builds on and extends it, adding specific details that are relevant to the programme, addressing issues related to the provision, exchange, availability, maintenance and use of data produced by MC-funded research activities or acquired from third parties as a support to this research. It thus applies to:

- Research projects funded under MC themes and cross-cutting activities following specific announcement of opportunity by NERC.
- Research funded by NERC through other funding mechanisms.
- Research undertaken by the MC Programme Directorate, except if funded by non-NERC sources. In this case other data policies are likely to apply. However, the MC Programme will still expect researchers to meet their within-project data management policy requirements and to consider making the data available for long-term curation.

The MC Data Policy aims to facilitate collaboration within and across the various projects, to protect the researchers' rights to publish their results within a reasonable time period (the right of 'first use'), and to further the dissemination, exploitation and scientific impact of results obtained by the programme's research projects.

To ensure the long-term availability of data generated by MC projects, NERC's Data Policy requires that the data sets assembled are offered to the most appropriate centre for the post-programme management. The Environmental Information Data Centre (EIDC), at the Centre for Ecology & Hydrology, will be the NERC data centre having primary responsibility for management of catchment data sets. As a NERC data centre, EIDC is committed to manage data for the long term, to allow post-programme exploitation. This commitment is given on the understanding that MC-supported projects will be responsible for the effective management and documentation of their data during the life of the project and that the EIDC has the right to refuse data which it considers are either of insufficient quality (for example, lacking appropriate documentation) or of little long-term value. All data deposited with the EIDC would be made publicly available via the CEH Information Gateway (<https://gateway.ceh.ac.uk>) after the 'right of first use' period has expired.

2 Types of data generated by the MC programme

Data sets to be generated by MC programme research fall into three categories: new measurements, model output and data synthesis.

1. The main sources of *new measurements* are likely to be:
 - a. Atmospheric concentration measurements for nutrients;
 - b. Measurements of nutrients in soils, river waters, groundwaters and estuaries;
 - c. Measurements of isotopes or continuous data from monitors;
 - d. Laboratory experimental data;
 - e. Biological and ecology data from field experiments and monitoring
2. *Model output* will be generated by new models of nutrient cycles, including new components to existing modelling systems (e.g. JULES).
3. *Data syntheses*, i.e. data sets derived from compilations, summaries or transformations of existing data that offer substantial added value over the original data, will be a key feature of the MC programme. New data syntheses will be developed in several fields, including catchment budget, fluxes to the sea and new analyses of earth observation data. There is a potential synergy between MC data synthesis and international initiatives of the International Geosphere – Biosphere Programme.

3 Data Management Planning

To ensure MC Programme data are appropriately managed and safeguarded for future long-term re-use, every funded project will be required to initially prepare, and then regularly maintain, its own Data Management Plan (DMP). The Data Management Plan should be a living document, regularly updated over the course of a project. Some of the key data management issues an MC programme-funded project needs to address (and, thus, should be covered in the full DMP) are outlined below. A checklist of issues is given in Appendix 2. Further information on DMPs will be included in the MC programme Data Management Guidance, which will be developed over the next few months.

4 Metadata

Production and maintenance of effective metadata are fundamental elements of data management activities during and after the project lifetimes. Metadata contain the supporting information to enable the effective utilisation of the data that they relate to. The quality of the metadata depends on their completeness and on how carefully they are selected, formulated and formatted. Good practice guidelines apply and will be detailed in the MC Data Management Guidance.

5 Data archiving, procurement, sharing and costs

Proposals will be expected to include outline plans for the management of data created, acquired, stored and used by the project. As part of this, PIs should identify the key data outputs of long-term value to the community, that they would submit for long-term storage by an appropriate NERC Environmental Data Centre. NERC operates a network of seven such Data Centres, four of which are relevant to the Macronutrient Cycles Programme:

- the Environmental Information Data Centre (EIDC), for freshwater (rivers and lakes) and terrestrial (e.g. land-use, soils) data;
- the British Atmospheric Data Centre (BADC), for data from atmospheric studies;
- the British Oceanographic Data Centre (BODC), for data from estuary studies; and
- the National Geosciences Data Centre (NGDC), for data from groundwater studies .

The Data Centres cannot take-on all data generated by projects and, therefore, projects will need to identify those data they consider should be retained to derive the maximum value to science into the future. For example, all the data from extensive laboratory experiments, or those from successive model iterations, would not usually be submitted for long-term storage. The NERC Data Centres would neither duplicate data stored elsewhere, such as those held by the National River Flow Archive, at CEH, the EA or the DEFRA DTC database.

The EIDC is the primary NERC data centre to be used by MC programme supported activities. To help researchers funded by the MC programme in undertaking data management, the EIDC will provide data management advice and support if required, **subject to a cost, which has to be built into any project proposal.**

Costs

It is recommend that the costs of data management be presented as two separate components in proposals. The first, should cover project participants' costs of managing data within the project itself and the preparation (conditioning) of data (of long-term value) for submission to the relevant Data Centre(s). The second component should cover the costs of the Data Centre(s). These are the costs the Data Centre would incur for liaising with projects, assessing long-term data storage requirements and providing support on data management, quality assurance, and the conditioning (formatting, metadata) of data that would be retained. Both components need to be built into projects because there will be no "top- slicing" of the MC Programme to cover data management. The costs of working with the Data Centres should be clearly indicated and be based on an estimate of the effort required and the Data Centre's daily payment rate. The Data Centres will not ordinarily charge for long-term storage costs, but may charge if the volume of the data exceeds 1TB. Applicants should contact the relevant Data Centre to establish likely costs. For successful projects, Data Centre costs will be a 'notional' part of the award, and will be provided directly to the Data Centres, rather than via the project, similar to Services and Facilities costs.

Storage and access to data

Data of long term value to the community must be lodged with the data centres (e.g. EIDC) as soon as they have been validated and no later than one year after acquisition. Data must be accompanied by such metadata as are required according to the MC Data Management Guidance. Responsibility for this rests with the PI of the individual MC-funded project, the MC programme fellowship holder, or the individual core team member. Exceptions to the one year deadline can be made in the case of model output data where after one year it cannot yet be determined whether the data will be suitable for long-term, post project curation.

It is particularly important to provide data in the required format for the EIDC, so that data can be easily stored in EIDC data management systems. Specialist assistance is available

from the EIDC for undertaking this data conditioning task but PIs must negotiate a cost for this work with the EIDC, and build these costs into their projects, as discussed above. However, it is generally envisaged that projects will possess the necessary data management expertise to enable data sets to be supplied to the EIDC in the correct format.

6 Archiving of model output

When archiving model output, archiving the supporting metadata is considered especially important. The metadata must contain all relevant information on the underlying theory (hydrology, chemistry, or statistics) as well as on the model configuration, input data, boundary conditions and the model code itself. Modelling groups are responsible for assuring a secure tracking system for model versions. All relevant metadata must be archived, although metadata and model output may be stored locally until the end of the relevant project. The exact procedure by which a decision on final archiving is made (by individual PIs, researchers, or collectively within the MC programme community) should be described in the project Data Management Plan.

7 Integration into the International Community

The MC programme's data management activities should make maximum use of existing international standards and international collaborations, provided this helps facilitate MC aims and objectives and does not run counter to the NERC Data Policy. MC programme data should be made available to support international programmes when data is requested and subject to the project PI taking responsibility for this international collaboration.

9 Access to MC data and Intellectual Property Rights

To facilitate data sharing, data will normally be available to all members of the MC programme community as soon as it has been submitted to EIDC. Within the first year after acquisition, use of the data by members of the MC programme community, other than the producers of the data, will require the prior consent of the producers. This is to allow individual PIs or co-investigators the opportunity for initial exploitation of the data they have created. However, all MC programme data creators are encouraged to support sharing of data within the MC programme community and are required to keep any 'embargo' period as short as possible. Any such period must not exceed 2 years from acquisition of the data. All MC data will become publicly accessible 2 years after acquisition. However, those who use the data within 2 years from the end of originating projects will be required to give the named originator(s) of the data the option of inclusion as a co-author on any resulting papers.

Enquiries about data storage and archiving can be addressed to Prof Paul Whitehead at paul.whitehead@ouce.ox.ac.uk or to the relevant data centre. In the case of catchment data the CEH Environmental Information Data Centre (EIDC) (http://www.ceh.ac.uk/sci_programmes/env_info.html) will be the NERC Data Centre with primary responsibility for the long-term management of MC catchment data sets.

Data collected as part of the project will need to be transferred to the EIDC in a timely manner and will need to be quality assured within each project. Applicants are advised

that their proposals must include the costs of managing and quality assuring their project's data; they must also meet the accession costs for the transfer of their data to the EIDC in an acceptable format.

Applicants may wish to contact the EIDC Data Management Coordinator, Dr Gwyn Rees (hgrees@ceh.ac.uk), to ensure their plans for data transfer are appropriate and the associated costs are properly accounted for.

APPENDIX 1

NERC DATA POLICY

NERC has a policy on data in order to:

- a. Ensure the continuing availability of environmental data of long-term value for research, teaching, and for wider exploitation for the public good, by individuals, government, business and other organisations.
- b. Support the integrity, transparency and openness of the research it supports.
- c. Help in the formal publication of data sets, as well as enabling the tracking of their usage to be tracked through citation and data licences.
- d. Meet relevant legislation and government guidance on the management and distribution of environmental information.

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. This includes data generated through complex systems, such as information retrieval algorithms, data assimilation techniques and the application of models.

This policy covers 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. This policy does not cover NERC's **information products**.

This policy will be reviewed at regular intervals to ensure it keeps pace with scientific requirements and data management best practice.

Key principles

The environmental data produced by the activities funded by NERC are considered a public good and they will be made openly available for others to use. NERC is committed to supporting long-term environmental data management to enable continuing access to these data.

NERC will supply the environmental data it holds for free, apart from a few special cases as detailed in the policy.

NERC requires that all environmental data of long-term value generated through NERC-funded activities must be submitted to NERC for long-term management and dissemination.

Access to data

It is NERC's policy that:

1. All the environmental data held by the NERC Environmental Data Centres will normally be made openly available to any person or any organisation who requests them.
2. The only restrictions on access which we will apply are those supported by the exceptions on disclosure in the Environmental Information Regulations (2004). If it is proposed to restrict access to any data we will explain why.
3. To protect the research process NERC will allow those who undertake NERC-funded work a period to work exclusively on, and publish the results of, the data they have collected. This period will normally be a maximum of two years from the end of data collection.

4. All data held by the NERC Environmental Data Centres will be supplied for free except for large or complex requests where we may charge the cost of supply, or where third-party licence conditions either prevent such free supply, or require us to make specific charges.
5. All environmental data made available by the NERC Environmental Data Centres will be accompanied by a data licence. Data originally provided to NERC by a third-party may have their own access and licence conditions which restrict how or when we can make data available to others, in which case our data licence conditions will reflect these.
6. All those who use data provided by NERC are required to acknowledge the source of the data.

NERC's Environmental Data Centres

Successful long-term data management requires both specialist data curation skills and an understanding of the science behind the data. NERC achieves this by supporting Environmental Data Centres and co-locating these within its research and collaborative centres to take advantage of the scientific expertise they possess.

It is NERC's policy that:

7. NERC will maintain Environmental Data Centres for the management and dissemination of environmental data of long-term value generated through NERC funding or deposited by third-parties.
8. The data centres will act impartially towards all data producers, regardless of whether they are based within or outside of NERC. The environmental data within the data centres will be open to all on the same basis.
9. Working with the environmental science community NERC will maintain criteria to identify environmental data of long-term value (a Data Value Checklist). These criteria will be used to inform all decisions that NERC makes on the acceptance and disposal of data by its data centres.
10. Information on all data held within the data centres will be made available through the NERC Data Discovery Service.

Data collection

NERC expects everyone that it funds to manage the data they produce in an effective manner for the lifetime of their project, and for these data to be made available for others to use with as few restrictions as possible, and in a timely manner.

It is NERC's policy that:

11. All applications for NERC funding must include an outline Data Management Plan, which must identify which of the data sets being produced are considered to be of long-term value, based on the criteria in NERC's Data Value Checklist. The funding application must also identify all resources needed to implement the Data Management Plan.
12. The outline data management plan will be evaluated as part of the standard NERC grant assessment process. All successful applications will be required to produce a detailed data management plan in conjunction with the appropriate NERC data centre.
13. All NERC-funded projects must work with the appropriate NERC data centre to implement the data management plan, ensuring that data of long-term value are

submitted to the data centre in an agreed format and accompanied by all necessary metadata.

14. Data from NERC-funded activities are provided to the data centres on a non-exclusive basis without prejudice to any intellectual property rights. This is to enable NERC to manage and make openly available publicly funded research data.
15. Those funded by NERC who do not meet these requirements risk having award payments withheld or becoming ineligible for future funding from NERC.

Open access to data underpinning research publications

NERC considers that long-term, open access to the data that underpin research publications will help to ensure the integrity, transparency and robustness of the research record. Access to these data supports the fundamental scientific requirement of allowing others to confirm or challenge research results.

It is NERC's policy that:

16. All research publications arising from NERC funding must include a statement on how the supporting data and any other relevant research materials can be accessed.
17. For all research publications produced by NERC's own staff, the supporting data will be made available through the NERC data centres.

* In line with UK Government policy, NERC distinguishes between **data** and **information products**. 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 human impacts 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 models. Whereas, **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. Model codes are not covered by this policy.

If you have any comments or questions, or require further information on the NERC Data Policy, please contact NERC's data management co-ordinator, Mark Thorley at mrt@nerc.ac.uk .

APPENDIX 2 MC PROGRAMME DATA MANAGEMENT PLANNING CHECKLIST

A checklist of issues that need to be considered when managing data in a MC Programme project.

Key: **Emboldened** = mandatory *Emboldened italicised* = mandatory if applicable Plain text = desirable/useful

Type of Information	Detail Needing to be Addressed
Basic Project Information	Project title , Project acronym
Project Organisation Relevant to Data	Project Leader , Data manager , Task leaders, Quality Manager, <i>External Collaborators</i> , <i>Subcontractors</i>
Project Data Management Timetable	Start date, major project activities (tasks) and dates , end date
Data Management Tasks & Responsibilities	Tasks/responsibilities not covered elsewhere (in the DMP, project documentation, etc.) e.g., field data input, validation etc.
Data Flow Summary	Major components of data use throughout project
Datasets	<ul style="list-style-type: none"> • Name, owner(s), storage location, estimated size, <i>3rd Party data</i>, <i>licensing arrangements</i>, • Access arrangements during project, • <i>If keeping data beyond project completion: date accessible, access conditions, download format through CEH Information Gateway or other. CEH Information Gateway Unique Record Identifier</i>
Quality Control of Data	What procedures will be put in place to ensure datasets are quality assured?
Data Back-up and Security	Arrangements for making data secure over the course of the project
Documentation/Methods	<ul style="list-style-type: none"> • How will datasets be created /acquired/ accessed? • <i>When will metadata be completed for the CEH Information Gateway?</i> • <i>Dataset summary including geographic range, file-naming, formats, content-level metadata (table&field), model /code documentation?</i> • Data storage and/or database design?

	<ul style="list-style-type: none"> • Where will documentation be stored?
Long Term Data Management Strategy	<ul style="list-style-type: none"> • Where will data be archived at the end of the project • If EIDC (CEH), who will be responsible for the data after project completion? • If another Data Centre, who will be responsible for transferring the data? • Will there be an embargo period required on access • Format of datasets, i.e., preferably not proprietary.
Data Management Plan	<ul style="list-style-type: none"> • Where will the DMP be stored? • When will the DMP be updated and reviewed?