# QUEST data policy

## 1 Aims and scope

QUEST aims to develop more realistic models and an improved predictive understanding of Earth system processes. Achieving this aim will mean using available observations more comprehensively and creatively than has been usual in Earth system modelling. Thus, the effective management of data generated by QUEST supported activities is an important priority.

The purpose of the QUEST Data Policy is to provide rules and guidelines to help researchers supported by QUEST to manage their data effectively in order to help them meet QUEST's scientific aims and NERC's Data Policy obligations. The QUEST Data Policy does not replace the NERC Data Policy. Rather, it builds on and extends the NERC Data Policy as provided in the NERC Data Policy Handbook (www.nerc.ac.uk/research/sites/data/policy.asp) adding specific details that are relevant to QUEST.

The QUEST data policy addresses issues related to the provision, exchange, availability, maintenance and use of data produced by QUEST funded research activities or acquired from third parties as a support to this research[1]. It thus applies to:

- Research projects funded under QUEST's themes 1, 2 and 3 and cross-cutting activities following specific announcement of opportunity by QUEST/NERC.
- Research funded by QUEST/NERC through other funding mechanisms.
- Research done by the QUEST core team and QUEST fellows, except if funded by non-NERC sources. In this case other data policies are likely to apply. However, the QUEST 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[2].

The Data Policy aims to facilitate collaboration within and across the various QUEST 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 QUEST funded research projects. Technical details on data handling and management, such as what data to consider for long-term curation, data formats, metadata creation and data storage, will be issued in a separate QUEST Data Management Plan following on from the scoping study undertaken to assess the data management requirements of the QUEST supported projects.

To ensure the long-term availability of data generated by QUEST projects, NERC's Data Policy requires that the data sets assembled by QUEST are offered to the most appropriate NERC data centre for long-term, post programme management. The British Atmospheric Data Centre (BADC, badc.nerc.ac.uk) will be the NERC data centre with primary responsibility for management of QUEST data sets. NERC gives a commitment to manage these data for the long-term, to enable long-term, post-programme exploitation. This commitment is given on the understanding that QUEST supported projects will be responsible for the effective management and documentation of their data during the life of the project and that the BADC has the right to refuse data which it

---

[1] Long-term, post project management of data acquired from third parties will be undertaken by NERC if IPR agreements allow.
[2] Long-term, post project management of QUEST data generated from non-NERC funded activities will be undertaken by NERC if IPR agreements allow.

considers are either of insufficient quality (for example, lacking appropriate documentation) or of little long-term value.

## 2 Types of data generated by QUEST

Data sets to be generated by QUEST 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 the global carbon cycle;
   b. Measurements on ice cores (e.g., CH4 isotopes).
   Other sources of new data are expected to be minor.

2. *Model output* will be generated by models of Earth system components (e.g. land biosphere, ocean, atmospheric chemistry) and by climate and Earth system models.

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 QUEST. New data syntheses will be developed in several fields, including palaeo-data and new analyses of earth observation data. There is a potential synergy between QUEST data synthesis and the Earth System Atlas, an international initiative of the International Geosphere – Biosphere Programme which has close links to QUEST.

## 3 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 QUEST Data Management Plan.

## 4 Data archiving, procurement and sharing

The BADC is the primary NERC data centre to be used by QUEST supported activities. To help QUEST funded researchers in undertaking data management, QUEST will be resourcing the BADC to provide within project data management advice and support.

Data must be lodged with the BADC as soon as they have been validated and no later than 3 months after acquisition. Data must be accompanied by such metadata as are required according to the QUEST Data Management Plan. Responsibility for this rests with the PI of the individual QUEST funded project, the QUEST fellowship holder, or the individual core team member. Exceptions to the 3 month deadline can be made in the case of model output data where after 3 months it cannot yet be determined whether the data will be suitable for long-term, post project curation (see Section 5).

Some of the other NERC data centres may also be involved in supporting QUEST if identified as opportune by the data management scoping study, however, BADC will remain as the primary data centre for QUEST. BADC will assist in the development of the QUEST Data Management Plan and it will also help facilitate access to data for researchers involved in QUEST projects to ensure

that unnecessary duplication of data procurement is avoided.  Sharing data within the community of QUEST projects will also be largely facilitated through BADC.

## 5 Archiving of model output

When archiving model output, archival of the supporting metadata is considered especially important. The metadata must contain all relevant information on the underlying theory (physics 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, including the model code, must be archived for all model runs that are used for research purposes (as opposed to pure testing). As an exception to Section 4 of this policy, the metadata and model output may be stored locally until the end of the relevant project.  In this case, the BADC must be informed and given access to the output as required.  This helps support the wider use of the output.  The long-term archiving of model output produced by QUEST research will be required to follow the BADC Policy and Guidelines for Archiving Simulations.  The exact procedure by which a decision on final archiving is made (by individual PIs, researchers, or collectively within the QUEST community) will be specified in the QUEST Data Management Plan.

## 6 Archiving of data syntheses

QUEST considers data syntheses a special case of model output.  Therefore, it is important that not only the data that were produced are archived, but also that the methods by which they were produced is thoroughly documented as part of the metadata.  If possible, the original source data (or references to it) should also be included in the archive.

## 7 Integration into international community

QUEST's data management activities should make maximum use of existing international standards and international collaborations, provided this helps facilitate QUEST aims and objectives and does not run counter to other elements of the QUEST Data Policy.  Specific recommendations for particular data types (see Section 2) are:

*New measurements*: Systems for quality assurance and public archiving of the types of data to be generated by QUEST have been developed at an international level by the carbon cycle and ice-core research communities. These should be used as much as is practical.

*Data syntheses*: QUEST is closely linked with the Earth System Atlas, an initiative of IGBP. The Atlas will function as a journal to which data sets will be submitted for publication.  It is hoped that this will provide the community of contributors with an incentive for data synthesis and documentation.  Data syntheses produced by QUEST researchers will be required to conform with standards set by the Earth System Atlas in collaboration with QUEST and the BADC.  If appropriate, synthesis data sets should be submitted to the Earth System Atlas for publication.

## 8 Access to QUEST data and Intellectual Property Rights

To facilitate data sharing data will normally be available to all members of the QUEST community as soon as it has been submitted to BADC. Within the first year after acquisition use of the data by members of the QUEST 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 QUEST data creators are encouraged to support sharing of data within the QUEST community and are required to keep any 'embargo' period as short as possible. Any such period must not exceed 1 year from acquisition of the data.

All QUEST data will become publicly accessible 1 year 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.

Restricted access will not apply to data created specifically for the Earth System Atlas. Once published in the Atlas, data will, by definition, become publicly available.

Where QUEST data sets have been deposited with the BADC, the IPR to those data need not be transferred to NERC. However, if the IPR are not transferred the data creators must grant NERC a non-exclusive licence to enable the reuse of the data to support *bona fide* research and public good activities. If requests for QUEST data are received for commercial use, NERC will refer them to the owners of the IPR unless other arrangements have been negotiated.

Any data sets that QUEST activities may create specifically for the Earth System Atlas will also be lodged at the BADC. However, they will be subject to more open access rules. This is because the Atlas is meant as a form of publication and is intended to be subject to the International Council for Science (ICSU) policy on free and open access to published data. Data in the Earth System Atlas will therefore be freely available as soon as they are published.

*Final version: Swindon, October 5th, 2006.*