

# Earth Science Academic Archive

## Data Management plan



The National Geoscience  
Data Centre



British Geological Survey, Keyworth, Nottingham, NG12 5GG

Introduction	3
The Earth Science Data Management Plan	4
Introduction	4
Responsibilities	4
Data provision and data acquisition	4
Metadata	5
Quality issues	5
Data formats	5
Data media	6
Data back-up policy	6
Data licensing issues/policy	6
Confidential or sensitive data	7
IPR issues/policy	7
Data Protocol	8
Longer term data management strategy	8
Other Legislation	9
Preparation of data management plans for Earth Science projects	10
Guidance on completing the Data Management Plan pro-forma	12
Project title	12
Introduction	12
Responsibilities	12
Data acquisition	13
Metadata	13
Quality issues	13
1.1.1 Verification	14
1.1.2 Validation	14
Data formats	14
Data media	15
Data back-up policy	15
Data licensing, copyright and general IPR issues/policy	15
Longer term data management strategy	16
Acknowledgements	17
Bibliography	17
Appendix A - Data Management Plan Pro-forma	18
Appendix B - Example of Data Management Plan	19
Appendix C - Example of a more complex Data Management Workflow	22

NERC Data Policy, as set out in the NERC Data Policy Handbook (NERC 2002) <http://www.nerc.ac.uk/data/documents/datahandbook.pdf> and requires that data management within the NERC Research Grant Programme be undertaken in a planned manner. One of the aims of this document is to formalise the implementation of the NERC Data Policy, for the specific needs of Earth Science Grants. It provides guidance and a framework that will enable appropriate data management. This document has borrowed freely from the documentation and experiences of data management in NERC thematic programmes

This plan is prepared at a high level and cannot address all the data management needs of individual parts of projects. It is one part of a data management hierarchy. For most small research grants only one plan will be required, but for more complex or involved research it should be supported by data management plans prepared for each element of the project. These data management plans should cover staff responsibilities, resourcing, data collection policies, data standards, data quality, quality assurance and IPR considerations, and policies on access to data, data dissemination and data marketing for each Project.

There is an obligation upon Earth Science PIs to ensure that data management is undertaken in a suitable way, and that adequate consideration is given to the "data side" of their work. To assist the PIs in meeting this obligation, this document provides guidelines on the preparation of individual data management plans, as well as providing a high-level data management plan for the Earth Science.

Inevitably, the detailed requirements of data management will vary from dataset to dataset, reflecting the type of data being managed, how the data are collected and stored, how the data might be used, and so on. Details of data management will be compiled and agreed at the most appropriate level, whilst conforming to broader guiding principles in place at higher levels within the Earth Science data management plan, which will in turn comply with guiding principles of best practice included in the NERC Data Policy.

## **Introduction**

The scope of this data management plan includes all activities within the Earth Science grant that result in output from the research including data collection or collation, data dissemination and data storage. All forms of output, including products such as models and software, are covered.

## **Responsibilities**

Responsibility for data at project level will lie with the appropriate PIs. PIs are expected to liaise with the Data Centre(s), whose staff will provide advice and guidance on data management and on items for inclusion in project-specific data plans. The Data Centre will produce and maintain their own data management plans.

For Earth Science overall responsibility for data management rests with the Data Centres who will provide advice and deal with routine data management, quality assurance and long-term strategy.

## **Data provision and data acquisition**

Negotiations for third-party datasets will be the responsibility of the PI for that project. The Data Centre can advise on the licensing/purchase of non-NERC data and they can undertake this on behalf of the scientist, to ensure the most favourable terms and conditions.

NERC-owned datasets will be made freely available to award holders via the Data Centres. Off-the-shelf digital datasets will be supplied free of charge to projects, unless additional data processing is required.

The Data Centres will collect and maintain information (metadata) about new datasets being generated by projects. Projects may wish to establish interim archives at an appropriate Data Centre during the life of the project. Data would then be secure and accessible for exchange with other projects.

Projects may choose to retain their own data until their work is concluded. However in the longer-term datasets will need to be deposited with an appropriate Data Centre, to ensure long-term stewardship and data access.

The NERC Data Policy requires projects to identify datasets for long-term stewardship by the NERC. Datasets derived from the NERC data should also be offered to the NERC under suitable terms. Guidance relating to submission of datasets to the Data Centres can be obtained from the Earth Sciences Data Manager

## ***In overall charge of Data Management***

Jeremy Giles  
Information Manager  
British Geological Survey  
Keyworth  
NOTTINGHAM NG12 5GG  
Tel: 0115 936 3100 Tel direct: 0115 936 3220  
Fax: 0115 936 3552  
E-mail: [jrag@bgs.ac.uk](mailto:jrag@bgs.ac.uk)

## ***Earth Science Projects Data Manager***

Rod Bowie  
National Geoscience Data Centre  
British Geological Survey  
Keyworth  
NOTTINGHAM NG12 5GG  
Tel: 0115 936 31 Tel direct: 0115 936 3106  
Fax: 0115 936 3276  
E-mail: rcb@bgs.ac.uk

## **Metadata**

Metadata - data describing the coverage, scope and derivation of a dataset - are important. The Data Centres will maintain appropriate metadata records for data holdings, whether of digital data, analogue data or materials. Wherever practical metadata will follow guidelines developed by the National Geospatial Data Framework (NGDF) for metadata provision, allowing that enhanced requirements may be desirable in some areas. The Targets for metadata provision should be specified within the data management plans. Technical Metadata may be required, for instance, to describe elements of the structure of in-house relational database tables, and to record any non-standard code dictionaries used in their population. When data are passed to Data Centres at the end of a project, the data owner must consult with the Data Centre manager to assess what level of any additional Detailed Metadata are required, to ensure the future accessibility and usefulness of the data.

## **Quality issues**

Earth Science projects are subject to Quality Assurance (QA), and each project's data management plan, and any approved data management procedures, will form part of the project's QA documentation. All projects must ensure that attention is paid to documenting the quality and reliability of all data, by providing appropriate metadata.

## **Data formats**

Wherever possible, unless project requirements dictate adoption of a specific format, digital data should be stored using standard, widely-available, software products and their related data formats. For example:

1. Oracle
2. SEGY
3. Arcview Grid Format
4. JPEG
5. TIFF
6. Microsoft Access
7. Microsoft Excel
8. Microsoft Word

The use of any other potential alternative, non-standard, software/formats should be discussed and agreed with the Data Centre Manager at an early stage. Discussions should include identification of appropriate strategies for the eventual transfer of the data and the safeguarding of their long-term accessibility and usefulness.

## **Data media**

Data media for acquisition of data from third parties and for data transfer between project teams and Data Centres should generally be:

1. CD ROM
2. DLT4
3. FTP

If CD ROMs are used, data should be transferred to secure media as soon as practical.

Use of any potential alternative media should be discussed and agreed with the Data Centre at an early stage.

Depending upon data volumes, direct digital data transfer (e.g. via e-mail or anonymous FTP) may be more appropriate than using intermediate media.

## **Data back-up policy**

The consequences of losing data due to having made insufficient or inappropriate provision for their back-up, are potentially catastrophic. Rigid periodic back-up programmes operated at the Data Centres safeguard major digital databases. Provision and support of back-up strategies for digital data stored locally is the responsibility of individual PIs, or their Researchers.

It is recognised that back-up of analogue data collections is less straightforward. Wherever possible, or financially viable, such collections should be “disaster proofed” by developing duplication strategies, such as micro-filming, digital scanning and photocopying and by ensuring the separate storage of “master” and second-generation copies.

All analogue data collections should be managed in accordance with appropriate National and International standards. For example, BS 5454:1989 provides a wealth of information on managing archival documents including paper, magnetic tape and optical disk; ISO 15489 Information and documentation – Records Management provides guidance on the management of all formation of data and MoReq the Model Requirements for the Management of Electronic Records.

PIs should bear in mind that the timely deposit of data with the Data Centres will provide additional security for the project data.

## **Data licensing issues/policy**

The Data Centres will facilitate access for projects to their data holdings, on the basis of appropriate end-use and onward dissemination conditions, and charging/licensing arrangements where appropriate. It is essential that all necessary licences are obtained before use of "third-party" data, equipment and applications. Compilers of data management plans will need to consider the local cost and functionality implications of entering into such "third-party" licensing agreements. The licences must also cover subsequent use by the Data Centre if required.

## **Confidential or sensitive data**

The PIs are responsible for ensuring that the terms and conditions for the use of any commercial, confidential or sensitive data are clearly established beforehand in writing. These conditions must allow for the eventual deposit of data in the Data Centre. The Data Centre will undertake to maintain any such data under confidential cover, if requested, but would insist that there are provisions for the release of the raw data at some reasonable time in the future.

## **IPR issues/policy**

Copyright is part of the wider field of Intellectual Property Rights (IPR), and is the intellectual property with which most PIs have some familiarity, via authorship of scientific papers. Where the results of research work are published in scientific journals and elsewhere (workshops or seminar proceedings or one-off special publications), copyright should be retained by the PIs, grant recipient organisations, or the NERC, and not assigned to the publishers. Such academic publishers are granted non-exclusive permission to publish the material or papers, thus ensuring that the copyright owners retain the right to reproduce the material elsewhere, and in whatever appropriate format.

Specific issues relating to the management and exploitation of IPR within the NERC are covered in the NERC guidance notes on the Exploitation of Data and Intellectual Property Management (NERC, 1996), which should be consulted and appropriate information/advice sought from the Institute contact.

In essence, IPR issues need to be addressed in all contractual arrangements, including those between the NERC and PIs, as well as those between PIs and collaborating organisations, prior to the start of a project. This ensures that the ownership rights of any intellectual property arising from that project, and future exploitation rights, are clearly defined and agreed between all potentially interested parties. This is important, even where no issues relating to exploitable IPR are expected to arise during a project.

Ownership of intellectual property and all intellectual property rights arising from NERC funding lies with the grant recipient organisation. In the case of NERC Centre/Surveys the ownership will normally remain with NERC as the legal entity, but the originating Institute is identified.

The NERC reserves the rights to retain, for a limited period, an exclusive right to exploit IP in partnership with grant recipient organisations, to the benefit of the UK and the grant recipient organisations. This is to avoid circumstances where fragmentation of IP (such as datasets) would

reduce the likelihood of successful exploitation. It is stated in the NERC policy that the IPR arising from NERC-funded work should be exploited, with the onus placed upon the research institutions to identify and exploit opportunities. Where appropriate, NERC, at an Institute level, will be directly involved in IPR protection and exploitation.

One of the aims of the Earth Science is to ensure that its results are made as widely available as possible. Whereas it is important to respect IPR ownership by PIs and/or their institutes for the duration of individual projects, it may be appropriate to limit the period of IPR protection beyond the end of individual projects, to encourage early exploitation. If no steps are taken to exploit IPR during this period, it may then be appropriate for the NERC to assume responsibility for exploitation. An appropriate period for IPR protection beyond the end of projects would be of the order of 1 year, but there may be justification for longer protection in some instances.

## Data protocol

The aims of this Data Protocol are

- to ease dissemination of scientific results within the community;
- to protect the rights of the individual scientists funded by ;
- to have all the involved PIs treated equitably;
- to ensure the quality of the data archive.

Although these aims may conflict at times, it is hoped that the provisions of the protocol resolve these conflicts fairly. It is recognised that this may not always be to everyone's complete satisfaction as there are bound to be cases where individual interests clash with those of the Earth Science as a whole. Therefore, to try to meet these aims, all scientists involved in the Earth Science are required to abide by the following conditions.

1. All scientists involved in are to have equal and complete access to measurements and model results produced during
2. Data must be made available to other scientists as soon as possible, *via* the Data Centres. Any corrections or amendments to data already archived at a Data Centre should be submitted as soon as possible.
3. To allow data validation, access to data and model results will be restricted to participants for a period of time ending one year after the completion date of the corresponding project. After the validation period, the data will be released to the public domain. Any extension of this period must be agreed with the Scientific Coordinator.
4. It is each researcher's responsibility to ensure that the data used in his/her publications are the best available at that time.
5. If measurements or model results from other research groups within are used in a publication, joint authorship must be offered. (This does not necessarily have to be accepted particularly in cases where due credit and acknowledgement can be given in other possibly more appropriate ways).
6. Whilst the data are restricted from the public domain, each researcher has the right to refuse to allow his/her work, whether measurement or calculation, to be used in a publication or presentation prior to the researcher's own publication of that work.

7. Whilst the data are restricted from the public domain, they should not be transferred to a third party (*i.e.* a non- participant) without the originator's consent.
8. Data and metadata should conform to the programme standards as documented. In the event of a dispute the final decision rests with the ESAA Manager.

## Longer term data management strategy

PIs have to offer the data resulting from their work to their Data Centre for long term archiving. PIs should liaise with their Data Centre throughout their projects to ensure that data transfer following project completion is undertaken in the most appropriate way. The Data Centre will be made aware of data types, formats and volumes prior to receipt, and can then advise on the most appropriate formats and media to be adopted. Strategies for the long-term management of lodged data need to be developed and agreed between the PIs and the Data Centres. The data must be lodged upon completion of projects, but may be sent at any appropriate time during the project, particularly if additional security is required. If data are retained by the project a strategy must be in place for their long-term preservation

It is important to recognise that data from the Earth Science may have a long-term value not evident at the time of their collection or generation.

The Data Centres will be responsible for the long-term management of the data lodged with them. In general their responsibilities will include:

- ongoing, potentially long-term, data management, once data have been collected;
- periodic review of all datasets held, and their classification as either:
  - essential (implying a need for provision of long term maintenance);
  - care and maintenance (data to be kept safe but not actively managed);
  - for disposal (no commitment for future data management);
- publicising any intention to put at risk or destroy data, before doing so, when it is considered that the cost of keeping datasets outweighs the apparent benefits. In such circumstances the Data Centres will take all reasonable steps to inform the relevant data providers of this intention, and will return data if requested.
- the projects PIs and their organisations will be responsible for any data retained by them after the end of the project. They must inform the Data Centres of any intention to put at risk or destroy data, before doing so, and provide the Data Centre with the data and appropriate metadata, if requested.

## Other legislation

The projects and PIs must ensure that they are aware of any current or planned legislation that may effect the provision of data or place obligations on them to retain data for future supply to PIs. For example the Environmental Information Regulations 1992, the Data Protection Act 1984 and the forthcoming legislation on Freedom of Information covering environmental data may have a direct effect on the supply of data to other PIs.

Data management plans are an essential part of the NERC Data Policy. A data management plan should be prepared for each new project. This plan sets out the data management framework for the project and identifies responsibilities and resource requirements.

It is intended that there will be a hierarchy of plans within the Earth Science below the over-arching Earth Science data management plan. These will be prepared to cover data management activities at Project level and subsidiary data management plans prepared for discrete parts of projects if appropriate. Depending upon the nature of individual projects, these subsidiary plans may only need to identify project-specific responsibilities etc., and refer to the higher-level data management plan on other issues. These higher-level plans should identify data management components that are common to many projects.

It is probable that the Earth Science data management plan will be sufficiently high-level that it will remain effectively static, though with scope for update or improvement if/when shortfalls are identified. In contrast, subsidiary data management plans are likely to be more dynamic and subject to mandatory periodic review and revision in the light of changing activities and technology. The status of Project data management plans will depend upon the life of the individual project and, for short-lived projects, planning for the long-term management of project data in a post-project situation must be considered at the outset, in line with parent NERC Data Centre policies.

It is not intended that the data management plans should be large, indigestible documents. They need to be understood easily by all staff and to identify responsibilities etc. clearly. They will form part of the Quality Assurance (QA) documentation of a project. Data management plans will be controlled documents, requiring rigorous version control and systematic distribution.

Data management plans will not normally define data management procedures. These are separate, detailed, data-specific, guidelines for the use and management of datasets. Whereas many procedures will be project/dataset-specific, forming part of the project's QA documentation, there are also likely to be a number that will be more generic, and may become formalised in procedures manuals.

Where appropriate, a data management plan should include elements such as:

- A defined data management infrastructure for the project;
- A preferred software (and perhaps hardware) specification;
- A statement on preferred data formats/media;
- A statement on QA policy;
- A statement on IPR protection/limitations, Copyright, etc.;
- A statement covering the use of any commercial or sensitive data

- A statement on the types of data that will be collected/acquired/created/maintained;
- A statement on data collection methodologies/policies;
- A statement on verification/validation methodologies;
- Appropriate standards for data accepted from (external) third parties or other Earth Science projects;
- "Disaster proofing" (at various levels);
- Treatment of data "acquired" but not immediately relevant to the Project activities;
- A recommended metadata standard (to cover Discovery, Detailed and Technical metadata, if appropriate);
- Consideration of data interchange within and outside the project, including plans for data dissemination;
- Long-term stewardship of the data.

In order to streamline the preparation of data management plans a pro-forma plan has been prepared (Appendix B). The pro-forma consists of a series of section headings to guide and standardise the preparation of plans. Not all sections will be relevant to all data management plans. However, in the majority of cases they will. Where sections are inappropriate in specific instances, suitable text should be inserted to indicate that the section was considered but found inapplicable during the preparation of the plan.

The data management plan should include headers and footers containing such information as project title, version and date, and should be page numbered using a numbering convention such as "Page 1 of 4".

Discussion of the contents of each section is provided below. Where "project" is referred to in this document it should be taken to read "project", "sub-project", "database" etc, as appropriate to the context of individual data management plans.

An example of the type of information required has been included (Appendix C) based on an existing NERC Thematic Project.

## **Project title**

An appropriate title for the area/topic covered by the data management plan. It should be concise and readily understood by all staff involved in the project (and many outside).

## **Introduction**

A brief introduction should be provided to set the context of the data management plan and the project for which it has been prepared. It may be useful to include information on where the project lies within the Earth Science and to identify its Data Centre(s).

## **Responsibilities**

A statement should be included to the effect that data should be managed to appropriate standards within the project, and the main purpose of the data management plan is to define these locally appropriate standards. They will differ between projects and, potentially, between different parts of the same project.

Practical responsibilities of individuals with respect to data management should be identified and defined. For responsibilities within a project individuals should be identified with specific roles. If responsibilities change with time a revised data management plan should be prepared and issued. Individuals identified should include, but not necessarily be limited to, Project Manager/Team Leader, Data(base) Manager(s) and their deputies.

If responsibilities outside the project (e.g. Intellectual Property Rights Manager) are identified, it is appropriate that they are referred to by role, rather than to identify specific individuals by name. Revision of the data management plan is then not required if individuals' responsibilities change.

An organogram of the project data management hierarchy should be provided

## Data acquisition

A statement on the data acquisition policy for the project should be provided. This should include information of what (and how?) data are being collected (and/or generated or derived), what they are being used for and how they will be managed. This information should include a statement on QA/QC of the data. If existing data are being used in a project from other or older (and pre-data management plan) sources, their compatibility with other project data should be assessed. The future "home" for such "imported" data also needs to be considered. For example, will these data form part of the project database. If they will, care must be taken to ensure that potentially non-concurrent data are not "duplicated" elsewhere.

## Metadata

It is essential that all data gatherers/generators provide appropriate metadata to their Data Centre, in line with current metadata standards, such as the "working standards" provided by Holmes et al, 1999. These "working standards" are in turn derived from more comprehensive NGDF and ISO standards.

The following definitions have been adopted by the BGS; it is proposed that they also be adopted for Earth Science use:

- **Metadata** are information that should allow a Data Manager to describe a dataset so that users can determine whether the data in the dataset will be of use to them.
- **Discovery metadata** are the minimum amount of information that is needed to describe a dataset so that users can determine whether the data in the dataset will be of use to them.
- **Detailed metadata** are data about the data themselves. Detailed metadata can be used to describe an individual row (or record), or parts of a row, as well as the dataset itself.
- **Technical metadata**, which are applicable to all subdivisions of data and underpin the structure of the other types of metadata, are concerned with how data are stored within database management systems.

## Quality issues

Data management is part of the Quality Assurance (QA) for a project and the data management plan forms part of the project's QA documentation. Advice on QA matters should be sought from the Quality Manager:

Rod Bowie  
National Geoscience Data Centre  
British Geological Survey  
Keyworth  
NOTTINGHAM NG12 5GG  
Tel: 0115 936 31 Tel direct: 0115 936 3106  
Fax: 0115 936 3200  
E-mail: rcb@bgs.ac.uk

Verification and validation of data are important processes that are often neglected and, sometimes, misunderstood due to their inconsistent usage in different situations. For clarity, brief descriptions and definitions that are useful in the data management context are provided below.

## 1.1.1 Verification

Data verification relates to the processes that establish the basic verity (truth) of the data, however this is assessed. Once this truth is established the reliability, quality, usefulness or fitness for purpose of the data may be measured, against fixed benchmarks or on a sliding scale, depending upon the uses to which they will be put. Wherever possible, responsibility for verification should lie with those who collect data and secondarily with those who use it. Verification can thus be carried out in two stages, the first before data are used, the second during their use, by comparison with adjacent data.

- **[Data] Verification** - the process or processes whereby a data collector, data user or data manager ensures that a data item [reading, measurement, observation, model, etc] is of appropriate quality, reliability and fitness for purpose.

## 1.1.2 Validation

Data validation is potentially a more specific and quantitative process, in which items of information are checked back against original (verified) data, including lists of valid, legal or legitimate values. Some data management systems include automatic validation checks that compare transcribed data with allowable codes or sensible values. The 'spell check' facility within word-processing packages is a simple example of validation software, which is limited by the supporting dictionary and its inability to recognise the use of correctly spelled words in inappropriate contexts. In systems that include conversion of analogue data into digital form, validation routines can be designed to deny entry of invalid information. However, no matter how sophisticated the validation system and regardless of whether it is manual or automatic, it cannot differentiate between data that are correct and incorrect, only between those that are valid or invalid.

- **[Data] Validation** - the manual or automatic process or processes whereby data users or data managers ensure that previously verified data are neither corrupted nor made invalid due to the introduction of accidental errors during transcription, digitisation or other forms of processing.

Responsibilities for data verification and validation of project data need to be defined. Reference to generic or project-specific data handling/management related procedures should be included. Project-specific procedures will be separate documents and not part of the data management plans, though by definition they will be part of the project's QA documentation. Quality labelling of data (part of the metadata) is important (Robson, 1994). The BGS guidelines on Discovery Metadata (Holmes et al, 1999) request information about the completeness, level of spatial detail, logical consistency and lineage (Lowe, 1995) of datasets.

## Data formats

Formats (including preferred software) to be used for the collection, manipulation, transfer and storage of project data should be specified. These will normally be common standards (such as

Oracle, Access, Excel, etc) that are readily portable. Obscure and obsolete formats and software should be avoided, unless there are over-riding project requirements for not doing so. In this event the reasons for the choice of format/software should be documented in the data management plan and appropriate provision should be made for the export of data in a readily portable form. The use of obscure and obsolete formats requires specific approval of the Data Centre.

## **Data media**

Preferred media for the transfer and storage of project data should be defined. The choice of media will depend upon individual project requirements, data volumes and the long-term management strategy for the data. Obscure and obsolete media should be avoided, unless there are over-riding project requirements for not doing so. If this is the case, the reasons for the choice of media should be documented in the data management plan and provision should also be made to ensure that data will remain accessible into the future.

## **Data back-up policy**

An appropriate project data back-up policy needs to be defined. This will be dependent upon the type/volume of data held, its “value” (in the broadest terms) and the longer-term data management strategy. It is important to consider the consequences to the project (and to the NERC) of losing data if data back-up provision is insufficient or inappropriate.

## **Data licensing, copyright and general IPR issues/policy**

There are two main aspects to consider in relation to data licensing. These are the third party interest (if any) in data being used in a project and the use of project data by third parties.

In the former case it is essential that there is compliance with the licence requirements of third party interests, for example Crown Copyright, data from other institutes/universities that may have “academic”-use strings attached, restricted use or commercial confidentiality issues. Where such restrictions exist (e.g. in the use of UK Hydrographic Office charts), they should be adhered to at all times, and identified within the data management plan.

In the latter case the commercial potential (not necessarily financial) and “strategic” value of project data may need the protection of licence agreements with any third party users of the data. The scope of such agreements needs to be considered.

Similarly, copyright and IPR issues relating to the data must be considered and terms and conditions agreed before the outset of any undertakings. It is particularly important in the case of co-funded and “commercial” projects to agree where copyright, IPR rights and exploitation rights reside at the outset of a project, and to provide formal written agreements. Such issues can be difficult or impossible (and potentially very expensive) to resolve later, if/when commercially “interesting” data arise from a project or agreements for third-party use of data.

Even if a project team does not think that these issues are important for a project, they may become so in the future, and they should always be addressed. There is a legal obligation to comply with the terms

and conditions of any third party owned IPR that are employed. Compliance is mandatory, not discretionary.

Seek advice on IPR and copyright rather than ignore the issue. In the first instance, guidance relating to any problems on Earth Science with these issues can be obtained from:

Chris Luton, Head of Intellectual Property Rights  
British Geological Survey, Keyworth, Nottingham, NG12 5GG.  
Tel: 0115 9363100; Tel direct: 0115 9363331; Fax: 0115 9363150  
E-mail: [chto@bgs.ac.uk](mailto:chto@bgs.ac.uk)

## **Longer term data management strategy**

It is essential that the data management strategy for project data beyond the end of a project be defined, especially for those items that will not be passed to a Data Centre. Failure to do so may result in the total or partial loss of data. This strategy needs to consider the “value” of the data in the future (or the consequences of its loss), the cost implications of the proposed strategy, contractual requirements for “commercial” work, responsibilities for implementation and periodic review of this strategy while the data are being managed. It should be remembered that almost all data are expensive to obtain and may be costly, difficult or impossible to re-collect in the future. Today's uninteresting data may well be highly important baseline data of the future.

The strategy adopted will range between “put it in a secure drawer and forget about it” through to formal deposition in centrally maintained data archives (e.g. a library or the NERC Data Centres etc). If the latter policy is proposed, responsibility for managing the data will effectively be passed on. The question of data donation must be raised with the intended recipients of the data at the earliest possible stage, so that the guidelines for the future use, archiving and ongoing management of the data can be discussed. In the case of digital data, active data maintenance may well progress through passive storage to eventual disposal.

The choice of media and software should be considered (will a CD written now be readable in 10, 50 or 100 years time?). This is less of a problem for paper records (though they may be microfiched or scanned), or for physical samples.

It is also useful to consider the future exploitation of project data. This is especially useful if commercial exploitation is intended (or even possible) following completion of the project, but should not be ignored because the data may be exploited in other ways (e.g. by another project) not yet anticipated.

Guidance relating to eventual submission of datasets to the NERC Data Centres can be obtained from:

Rod Bowie,  
National Geoscience Data Centre  
British Geological Survey  
Keyworth  
NOTTINGHAM NG12 5GG  
Tel: 0115 93631; Tel direct: 0115 9363106; Fax: 0115 9363276  
E-mail: [rcb@bgs.ac.uk](mailto:rcb@bgs.ac.uk)

## Acknowledgements

This document was adapted from the data management documentation of the URGENT Programme and built upon the experience of members of the URGENT DMQA Committee. The documentation drew on the experience and input of many NERC institute staff members who are thanked for their valuable input.

## Bibliography

British Standard 5454:1989. Storage and exhibition of archival documents. British Standard Institute.

International Standard ISO 15489-1:2001 Information and documentation - Records Management

International Standard PD ISO/TR 15489-2:2001 Information and documentation - Records Management - Guidelines

European Commission, 2001. MoREQ Model Requirements for the management of Electronic Records

Holmes, KA, Dobinson, A, Giles, J RA, Johnson, C C, Lawrence, D J D and McInnes, J L, 1999. BGS-geolDS Metadata - Issues Document. British Geological Survey Technical Report, WO/99/01R.

ISO Draft Standard, 1997. Geographic information - 15046 Part 15: Metadata. ISO/TC 211/WG3/PT 15/N0111.

Lowe, D J, 1995. The geological data manager: an expanding role to fill a rapidly growing need. 81-90 in Giles, J RA (editor), Geological Data Management. Geological Society Special Publication No.97.

NERC, 1996. Exploitation of Data and IPR

NGDF Discovery Metadata Guidelines, 2000. Version 1.2b.

Robson, P G, 1994. Quality Assurance of Data in the BGS. British Geological Survey Technical Report, WO/94/11R.

NERC, 1999. NERC Data Policy Handbook v 2.1

NERC, 2000. The URGENT Programme Data Management Plan

### Data Management Plan Pro-forma

XYZ Project Data Management Plan

---

Project Title

Introduction

Responsibilities

Project Manager:  
Data Manager:

Data acquisition

Metadata

Quality issues

Data formats

Data media

Data back-up policy

Data licensing issues/policy

Data copyright issues/policy

Data IPR issues/policy

Longer term data management strategy

Appendix 1- List of component [lower level] data management plans

Prepared by:

Date:

Checked by:

Date:

Approved by:

Date:

## Example of Data Management Plan

*Please Note: Items in Italics indicate details that need to be completed or are examples of the type of information required. This information does not need to be too detailed but should provide sufficient to provide an overall picture of what is involved in the Project's data management*

### Project Information

Project: <b>(RG1234)</b>	<i>Project title</i>
Project reference:	<i>NERC research grant reference number</i>
Principal Investigator(s):	<i>Name(s) and full details, i.e. address etc., of PI(s) with contractual responsibility</i>

### Project Organisation

Person with main responsibility for data(s):	<i>Name of the person(s) having overall responsibility for management of project data. Include contact details if different from above</i>
Nominated Data Centre: <i>For example:</i>	<i>Rod Bowie Records Officer National Geoscience Data Centre British Geological Survey Keyworth Nottingham NG12 5GG</i>
	<i>Telephone:+44(0)115 9363106 Direct/Pager Fax:+44(0)115 9363276 Office e-mail: <a href="mailto:rcb@bgs.ac.uk">rcb@bgs.ac.uk</a></i>
Other Team Members :	<i>If appropriate add names, titles and contact details of other team members (technical and administrative) including a description of duties.</i>

## 1. Introduction

This Plan covers staff responsibilities, data collection, data standards, data quality, and IPR considerations for *(insert project number)*, entitled "*(insert project title)*".

## 2. Responsibilities for management of the project data

*(include any relevant tasks or additional responsibilities)*

Principal Investigators: e.g. Purchase of data from 3rd parties

Researcher: e.g. *Capture of data from the field*

*Production of metadata*

*Transfer of metadata and data to the Designated Data Centre*

### 3. Data acquisition

List all the different datasets that will be collected or generated by the project, together with a brief description of each dataset. This includes datasets that were supplied to you by a third party.

*For example:*

*Three types of data: (a) lithological, (b) geophysical and hydraulic, pertaining to the Triassic sandstone aquifer in Birmingham and Nottingham will be acquired under this project.*

*a. Lithological data will be collected through observations recorded in a dated and numbered field notebook during the procedure of test drilling. This information will be copied into a digital format (Excel spreadsheets) by the Researcher.*

*b. Geophysical data will be recorded on A4 paper by the School of Earth Sciences, University of the North of England (subcontractor), during the procedure of geophysically logging test boreholes. This information will also be provided in digital format (Excel spreadsheets).*

*c. Hydraulic data will be collected both digitally by datalogger and manually by the Researcher through observations recorded in a dated and numbered field logbook during the procedure of hydraulic testing of boreholes. This information will be copied into a digital format (Excel spreadsheets).*

### 4. Metadata

You will need to provide sufficient metadata with your data sets to ensure that they can be understood and used by other researchers. Insert a statement to indicate how the metadata will be supplied.

*For example: Discovery metadata for all datasets will be included and recorded as “headers” in each Excel spreadsheet used to store the acquired data.*

### 5. Quality issues

List the procedures and provide references if appropriate for the quality control carried out on your datasets.

*For example: Internal calibration (for geophysical data), instrument calibrations, duplicate samples and field blanks (for hydrochemical data) will be recorded and tested against collected/recorded data to ensure their validity. Qualitative descriptions (lithological data) will be validated through comparative descriptions of collected materials.*

### 6. Data formats

List the formats in which the datasets will be provided to the Data Centre.

*For example: All data collected by the project will be supplied in Microsoft Excel spreadsheets.*

*Statistical and model outputs will be submitted in the following formats together with model descriptions: Surfer (dxf format) Visual Modflow (vmf format)*

### 7. Data media

List the media on which the datasets will be stored and later provided to the Data Centre.

*For example: All data will be stored on the hard disk of desktop and laptop computers. Long-term storage of collected data will employ compact disks and the University SAN and archive facilities*

*Data will be supplied to the Designated Data Centre on CD-ROM or via ftp*

### 8. Data back-up policy

Specify your back-up procedures.

*For example: Recognising the susceptibility of hard disks to failure, collected digital data will be transferred on a weekly basis to the University SAN*

*As soon as each discrete element of research is completed the data will be copied to the Designated Data Centre with accompanying metadata.*

### 9. Data licensing issues/policy

List any licence issues relating to the data sets that you have collected.

*For example: The main output from this project is field data. We recognise that these data are the property of NERC and hence we will be asking their permission to licence these data to XYZ Petroleum for use in their exploration programme.*

### 10. Data copyright / IPR issues/policy

List any copyright / IPR / policy issues relating to the data sets that you have collected.

*For example: NERC and third party copyright will be protected. The PI will be responsible for ensuring that all project members are aware as to the ownership of data and who may access them and under what conditions. On-line access to the data will be password protected.*

### 11. Longer term data management strategy

Document your strategy for the long-term security of your datasets, techniques and/or software.

*For example: All data collected under this project will be stored electronically (compact disk) as a final report appendix. The final report will be placed in a downloadable format on the Internet websites of the University of the North of England. Data will also be submitted to the NERC Designated Data Centre at the British Geological Survey*

Prepared by:

Name:

Date:

Checked by:

Name:

Date:

Approved by:

Name:

Date:

### Example of a more complex Data Management Workflow

#### Ocean Margins Link Data Management Workflow

