

# UKCCSRC Data Archive

# 2015

## Guidelines for Depositors



The UKCCSRC Data  
and Information Life Cycle



Rod Bowie, Maxine Akhurst,

Mary Mowat

UKCCSRC Data Archive

November 2015

# Table of Contents

<i>Table of Contents</i>	2
<b>Preface</b>	1
<b>About the UKCCS Data Archive</b>	2
<b>The UKCCSRC Data and Information Life Cycle</b>	4
<b>Step 1 – Proposal Planning and Writing</b>	5
<b>Step 2 – Project Start-up and Data Management</b>	6
<b>Step 3 – Dataset Creation</b>	7
<b>Step 4 – Analysing and Maintaining Your Data</b>	8
<b>Step 5 – Preparing Data for the UKCCSRC Data Archive</b>	9
<b>Step 6 – Depositing the Results of your Research</b>	11
<b>Step 7 – Post Deposit Activities</b>	12
<b>UKCCSRC Data Archive Contact Details</b>	13

# Preface

## *Guidelines for preparing, donating and depositing data*

The UKCCSRC funds projects to promote the implementation of Carbon Capture and Storage (CCS) and require data from these projects to be lodged in the UKCCSRC Data and Information Archive. Natural Environment Research Council (NERC), the Engineering and Physical Sciences Research Council (EPSRC) and other research funders also provide grants for CCS scientific research. All research council grant holders have to abide by the conditions of the grant and follow their research council expectations such as the NERC Data Policy and EPSRC Research Framework. These policies place a duty on grant holders to offer a copy of their datasets to the appropriate Data Centre. The UKCCSRC Data and Information Archive is a component of the National Geoscience Data Centre<sup>1</sup> (NGDC), hosted within the British Geological Survey (BGS).

CCS data collected by UKCCSRC and EPSRC-funded projects must be deposited in the UKCCSRC Data Archive in addition to any home institution requirements. CCS data from projects supported by NERC and other funding streams can also be deposited with the UKCCSRC Data Archive, on a voluntary basis, under the same terms and conditions as a UKCCSRC-funded research.

This document explains the benefits that this deposit offers, describes the process of depositing information, and provides information about other sources of relevant information.

Two sources were drawn on during the preparation of this document. The first is the Research Information Network's document entitled – "Stewardship of digital research data - principles and guidelines"<sup>2</sup>. The second document is the ICPSR Guide to Social Science Data Preparation and Archiving 3<sup>rd</sup> Edition 2005. We gratefully acknowledge the contribution that both these documents have made to these guidelines.

---

<sup>1</sup> <http://ngdc.nerc.ac.uk/services/ngdc/>

<sup>2</sup> [www.rin.ac.uk/data-principles](http://www.rin.ac.uk/data-principles)

## About the UKCCS Data Archive

The UKCCSRC recognises the importance of maintaining a major national collection of CCS research information in an established archive and selected the National Geoscience Data Centre (NGDC), established in 1984, to host the UKCCSRC Data and Information Archive. The NGDC manages data and information created by the academic community, central government and many commercial sectors. The UKCCSRC Data and Information Archive holds data collected by all disciplines engaged in CCS research.

The UKCCSRC Data Archive component of the NGDC is responsible for working with UKCCSRC, EPSRC and other CCS grant holders to ensure the long-term preservation, promotion, discovery, re-use and re-purposing of their data. All creators and custodians of analogue and digital CCS data created by grant holders are expected to deposit a copy of the data with the UKCCSRC Data Archive.

The UKCCSRC Data Archive seeks to manage information in accordance with EPSRC and NERC data policies and will:

- manage information of all formats in appropriate environments and use appropriate technologies to ensure its long-term preservation and availability;
- promote the datasets managed with the UKCCSRC Data Archive via the UKCCSRC Data Archive website;
- ensure that appropriate metadata has been provided by the scientists who deposited the data and that this metadata is disseminated to aid discovery of the resource;
- maintain a system in which all records, in whatever format or media, can be found rapidly;
- encourage the re-use and re-purposing of the data deposited with the UKCCSRC Data Archive in future scientific research.

The UKCCSRC Data Archive supports the Research Information Network<sup>3</sup> (RIN) principles for “Stewardship of digital research data<sup>4</sup>”. These RIN principles state:

*“In order to produce high-quality research, researchers must have access to as wide a range as possible of the data and information produced by other researchers, as well as relevant information produced by other agencies in the UK and overseas. Similarly, successful dissemination and exploitation of research depends on effective flows of information between researchers and other individuals and organisations that have an interest in its results. A successful research and innovation system thus depends on the open exchange of ideas, information and knowledge.*

*Developments in information and communications technologies are transforming the nature and scale of research, enhancing both quality and productivity. They are facilitating new kinds of research, new organisational models, and collaboration across disciplinary, institutional and national boundaries. But they also demand new ways of thinking about how we manage data and information outputs, so that we can maximise their value, and ensure that precious resources are not lost. In pursuance of those goals, the fundamental policy objective is to ensure that*

***Ideas and knowledge derived from publicly-funded research should be made***

---

<sup>3</sup> [www.rin.ac.uk](http://www.rin.ac.uk)

<sup>4</sup> [www.rin.ac.uk/data-principles](http://www.rin.ac.uk/data-principles)

**available and accessible for public use, interrogation, and scrutiny, as widely, rapidly and effectively as practicable.**

*To help achieve that objective, the following five principles provide a broad framework for developing good practice for universities, research institutions, libraries and other information providers, publishers, research funders as well as researchers themselves. The principles are pitched at a high level. The guidance set out in the main document provides pointers as to how policy and practice may need to be changed to ensure that they are to comply with the principles*

- I. The roles and responsibilities of researchers, research institutions and funders should be defined as clearly as possible, and they should collaboratively establish a framework of codes of practice to ensure that creators and users of research data are aware of and fulfil their responsibilities in accordance with these principles.*
- II. Digital research data should be created and collected in accordance with applicable international standards, and the processes for selecting those to be made available to others should include proper quality assurance.*
- III. Digital research data should be easy to find, and access should be provided in an environment which maximises ease of use; provides credit for and protects the rights of those who have gathered or created data; and protects the rights of those who have legitimate interests in how data are made accessible and used.*
- IV. The models and mechanisms for managing and providing access to digital research data must be both efficient and cost-effective in the use of public and other funds.*
- V. Digital research data of long-term value arising from current and future research should be preserved and remain accessible for current and future generations.”*

UKCCS Data Archive staff are happy to provide advice and support during the course of the project to ensure that the data is in good order and well documented at the end of the project.

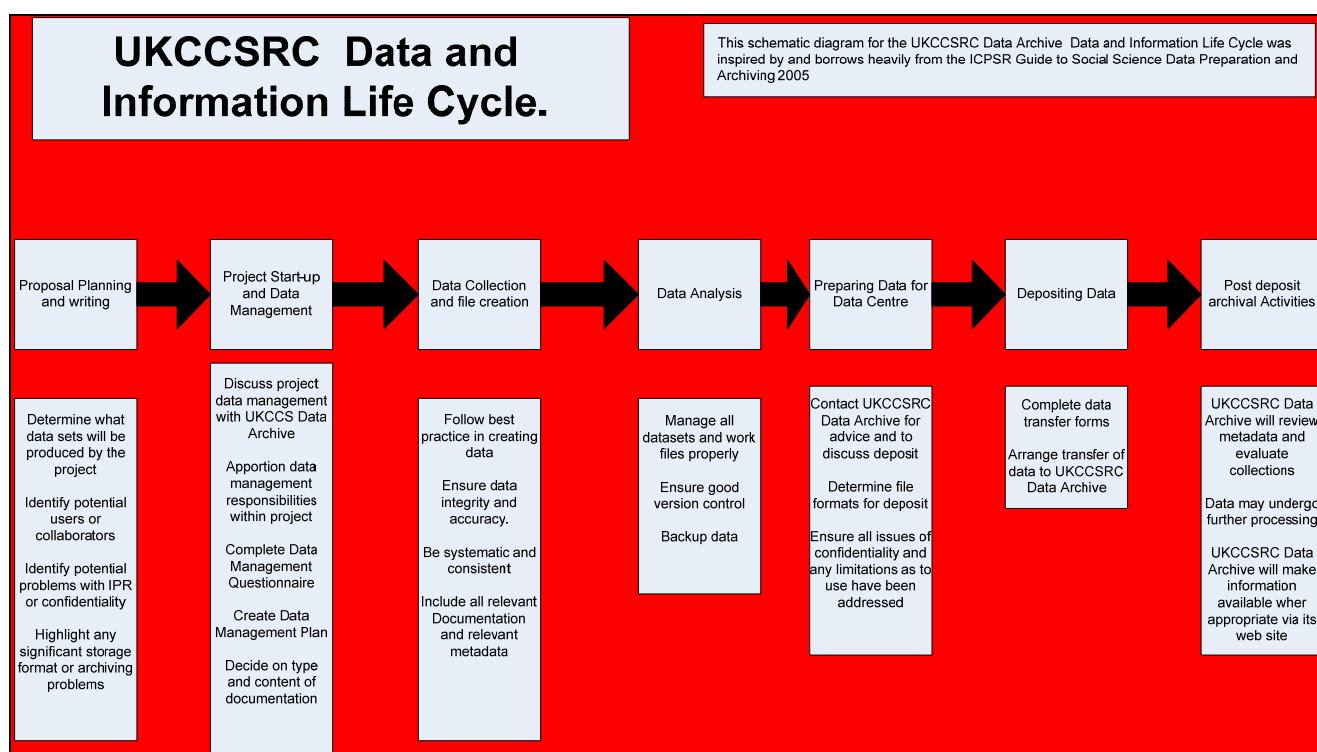
# The UKCCSRC Data and Information Life Cycle

## Using these guidelines

This document is aimed at supporting depositors with the UKCCSRC Data Archive engaged in any part of their project, from applying for a research grant, through the data collection, and to the preparation of the data for deposit in the archive. Researchers can contact the UKCCSRC Data Archive at any time to discuss their plans with respect to the selection and preparation of datasets.

Many investigators are more than willing to make their data available to others, but are unsure of how to go about preparing data for outside use, particularly in terms of complete documentation. The guidelines are intended to help researchers document their datasets and prepare them for deposit with the UKCCSRC Data Archive.

The main steps are set out below, in an order that reflects the lifecycle through which research data are created, used and made accessible over the long term.



# Step 1 – Proposal Planning and Writing

It is important to contact UKCCSRC Data Archive staff as soon as possible. They will provide advice and assistance on the following:

- datasets that already exist that may be of use to the research;
- data management planning for the grant proposal;
- how to document the datasets created and what metadata will be required;
- advise on periods of exclusive use by the creators of the dataset;
- what would be appropriate for final deposit with the UKCCSRC Data Archive; and
- costs to include in the grant proposal for data management and archiving.

If you are developing a proposal for a research grant then the proper management of the data is considered to be part of the project and failure to deal adequately with this aspect may contribute to rejection or cause delay in approval. By contacting the UKCCSRC Data Archive you can gain professional advice and guidance that will help you meet the standards expected in grant applications.

It is important to consider Intellectual Property Rights (IPR) and copyright at an early stage in proposal planning and writing. The eventual aim is to ensure that datasets created as part of the research will be available to for re-use and re-purposing by future researchers. To this end researchers are asked to specify that the UKCCSRC Data Archive should have a non-exclusive license to distribute the deposited datasets. The UKCCSRC Data Archive can agree reasonable embargo periods.

Where researchers are planning to use commercial datasets, owned by third-parties, they are encouraged to set up agreements so that the data can be released. The UKCCSRC Data Archive may store a copy of the dataset on a Commercial-in-Confidence basis for an agreed period. The UKCCSRC Data Archive will acknowledge that the copyright of these datasets are vested with the organisations that created them.

# Step 2 – Project Start-up and Data Management

It is important to contact UKCCSRC Data Archive staff once the funding has been agreed so that detailed data management planning my progress.

The first step will be ensuring that all parties are clear about their roles and responsibilities and their expectation of other parties that are involved. This is in line with first of the RIN principles for the stewardship of research data which states that:

*The roles and responsibilities of researchers, research institutions and funders should be defined as clearly as possible, and that they should collaboratively establish a framework of codes of practice to ensure that creators and users of research data are aware of and fulfil their responsibilities in accordance with the principles set out in this document.*

Planning for the management and storage of data at the beginning of the project is essential if the data is to be well managed throughout the life of the project and archived effectively. Without adequate data management planning data can be lost or corrupted and the costs of archiving increased.

The types of project documentation that will be required need to be considered. How will key decisions be documented during the projects life? What standards will be adopted by the project? How will these standards be implemented? What procedures need to be developed and how will they document, disseminated and preserved? What ontologies will be adopted or created and how will these be managed and preserved? How will change be managed?

Datasets should be well documented at the time of collections. This requires early planning to make sure that consistent documentation is created. Standards are now emerging for data product specification<sup>5</sup> and these should be considered in the process of planning the documenting of datasets. Appropriate metadata will be required at the time of deposit to aid discovery.

---

<sup>5</sup> ISO 19131:2007 Geographic information – Data product specification



# Step 3 – Dataset Creation

The second RIN Principle states:

*Research data should be created and collected in accordance with applicable international standards, and the processes for selecting those to be made available to others should include proper quality assurance.*

“From the archivist’s and end user’s perspective, a “good” dataset is one that is easy to use. Its documentation is clear and easy to understand, the data contains no surprises, and users are able to access the dataset with relatively little start-up time.<sup>6</sup>” UKCCSRC Data Archive staff would add that it has sufficiently detailed metadata to ensure it could be easily discovered. To achieve these goals effective data management must begin early.

It should be assumed that datasets deposited with the UKCCSRC Data Archive will be seen and re-used by other researchers. It is therefore important to adopt appropriate international standards and be able to demonstrate the use of best practice, so that those re-using and re-purposing the information have confidence in the quality of the datasets. This confidence is demonstrated by having a clear understanding of the authenticity, reliability, lineage and completeness of the dataset. This in turn is built by having clear documentation of their origin, including the methods, standards, procedures, instruments and techniques used in their creation or collection?

Where datasets are created in collaboration with other researchers it is essential that the IPR and copyrights issues are clearly defined.

During the course of the project appropriate security measures must be in place. For digital data appropriate measures should be used to protect the datasets from virus, theft of equipment, etc. A backup regime should be in place to ensure that datasets and supporting documentation are protected. Analogue information and specimens should also be protected appropriately.

---

<sup>6</sup> See page 7 of ICPSR Guide to Social Science Data Preparation and Archiving 3<sup>rd</sup> Edition 2005

# Step 4 – Analysing and Maintaining Your Data

The third RIN Principle states:

*Research data should be easy to find, and access should be provided in an environment which maximises ease of use; provides credit for and protects the rights of those who have gathered or created data; and protects the rights of those who have legitimate interests in how data are made accessible and used.*

The research community and anyone with an interest in the data should have timely, user-friendly access. This does not mean that all data should be made available immediately to all who may have an interest in it. There is a need to balance conflicting interests and rights. The requirement is for access in a managed environment; and to achieve that, a number of issues must be addressed.

The ability of users to find and re-use data depends on having a full record of information relating to the content, structure, context and source of data that may be relevant to their needs. Such information – data about data – is commonly referred to as metadata. Where appropriate, metadata should be created and made available in accordance with recognised international standards. ***But the essential requirement is as full a record as possible of the context in which the data was created or collected.*** Without this information, users cannot be aware of the nature and limitations of the data and will not be able to interpret them properly, or use them effectively.

Data are of no use without the facilities, software applications and other tools required to access and use them. Most users will have access to these applications and tools; but use of the data may only be possible if steps are taken to make the necessary applications and tools available.

All relevant files, particularly datasets under construction, should be backed up frequently to prevent having to re-enter data. It is a good practice to maintain a master version of the dataset that is stored on a “read only” basis. These master datasets should be backed up every time they are changed in any way. Computing environments in universities and research centres support data backup and storage. Although everyone knows the importance of backing up data, the problem is that few actually follow this through. It is also advisable to maintain a backup copy of the data off-site. One fire can destroy many years of work.

Physical specimens and samples are a vital part of many projects and they should be well labelled, properly documented and kept securely. Doing sophisticated analyses or producing complex theories based on poorly recorded specimens is obviously of little value. The UKCCSRC Data Archive can provide a secure repository for materials at any time.

Data should be open if possible. If this is not the case (3<sup>rd</sup> party data are included) there should be a clear process for determining who should have access to what data, and on what terms.

Copyright and the ownership of datasets and work must be considered and researchers should ensure that they have a right to transfer any data received from other parties or used as part of their research. This means establishing relevant procedures and audit trails. Any agreements should be made clear to UKCCSRC Data Archive staff who can provide help and advice.

Users will want to access research data alongside publications, in order to assess the evidence on which the research is based. It is important that researchers should receive appropriate credit for the work that has gone into creating or collecting data. Citation of research data (alongside citation of publications) is important in evaluating research quality.

# Step 5 – Preparing Data for the UKCCSRC Data Archive

## Evaluating data for permanent retention

The fifth RIN Principle states:

*Research data of long term value arising from current and future research should be preserved and remain accessible for current and future generations.*

Not all research data are of long-term value. Many datasets may have little value beyond the life of a specific project. It is essential that researchers put effective procedures in place to determine which data are of continuing value, and to ensure that arrangements for their stewardship are sustainable. Some datasets are of such significance and value that they need to be preserved and made accessible over the very long-term.

As research data are increasingly updated, amended and annotated over their lifecycle both by data creators and by subsequent users, provenance protocols and audit trails are needed to indicate clearly who has annotated or amended data, how and when. Otherwise there is the danger that data will be misinterpreted or used inappropriately.

### What needs to be done?

Ideally preparation for the eventual deposit of data should begin at the start of the project. This may initially be best done by establishing a dialogue with the UKCCSRC Data Archive Collection Officer. Preparation of Data Management Plans and properly documenting the information and metadata is essential.

Any documentation should enable a third party to make sense of the data. The UKCCSRC Data Archive data management questionnaire is a good starting point. This makes donors think about the data and information outputs the project will create. However, there may be documentation requirements specific to particular types of data or research, so potential depositors should ensure these are recorded. Much of the required basic data will be held or supplied in the accompanying metadata, but additional information helps with the understanding of the data. The following list provides some headings for additional information that could be of value:

- **Background to the project**
- **The purpose of the project**
- **Information about methodologies**
- **Sources used**
- **Related research**
- **Sampling procedures**
- **Content and structure of dataset**
- **List of filenames and description of contents**
- **Any thesaurus or dictionaries used**
- **Explanations of codes or anagrams used**
- **Description of any known errors or weakness in the data**
- **Documentation of any record or data conversion or format change**
- **References to publications resulting from or related to the project**
- **List any information about other Data Centres, Universities, etc. that may hold material related to the dataset**

- **Indicate of how long archive is to be retained (indefinitely or for fixed period)**

# Step 6 – Depositing the Results of your Research

Potential donators should contact the UKCCSRC Data Archive Data Collection Officer for information about deposit either by e-mail or telephone (see Contacts). Researchers are advised to complete the UKCCSRC Data Archive questionnaire well in advance which will cover basic details and covers frequently asked questions.

Datasets should be accompanied by metadata (information about the data), which should describe the datasets and stipulate how and by whom the data can be reused. The Metadata form is available for download or from the Data Collection Officer.

Datasets will be assessed by UKCCSRC Data Archive staff to ensure they are appropriate for deposit with the UKCCSRC Data Archive.

Once accepted, datasets will be accessioned using the standard UKCCSRC Data Archive workflow. Accessioning procedures include data validation, indexing and databasing.

# Step 7 – Post Deposit Activities

The fourth RIN Principle states:

*The models and mechanisms for managing and providing access to digital research data must be both efficient and cost-effective in the use of public and other funds.*

To maximise the benefits and minimise the costs of managing and research data, it is essential that research data must be both efficient and cost-effective in the use of public and other funds.

Although it is costly to maintain this information, managing and providing access to research data brings huge potential benefits in enhancing the overall efficiency of research; avoiding unnecessary duplication of effort and needless costs in knowledge transfer. Effective and efficient management of data requires specialist professional support services, which are provided by the UKCCSRC Data Archive to ensure that data is properly selected and stored, that it can readily be accessed, and that its integrity can be ensured over time.

The UKCCSRC Data Archive undertakes to:

- To maximize the benefits of research data information deposited with it and to make this information readily available.
- Ensure that suitable metadata is available for all the information deposited.
- Manage the information provided, in whatever format, in an effective way and use appropriate technologies to ensure its long-term preservation and availability.
- Create a system in which all data, in whatever format, can be found readily.

# UKCCSRC Data Archive Contact Details

The Manager  
UKCCSRC Data Archive,  
British Geological Survey,  
Keyworth,  
Nottingham,  
NG12 5GG  
tel: 0115 9363106  
[ukccsrc\\_data@bgs.ac.uk](mailto:ukccsrc_data@bgs.ac.uk)

Data Collection Officer  
UKCCSRC Data Archive,  
British Geological Survey,  
Murchison House,  
West Mains Road  
Edinburgh,  
EH9 3LA  
tel: 0131 6500312  
[ukccsrc\\_data@bgs.ac.uk](mailto:ukccsrc_data@bgs.ac.uk)