

# Data management in UPGro

Rod Bowie April 2015

# Take data management seriously

---

Please take three things away today - they are:

- That you should take data management seriously
- That data management is not a chore, if it is done at the right time
- That the Data Centre is there to help.

# Principle

- 
- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.

# NERC data policy

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

# NERC data policy

- 
- Environmental data - individual items or records (both digital and analogue) obtained by measurement, observation or modelling of the natural world, including calibration and quality control. Includes data generated through complex systems, such as information retrieval algorithms, data assimilation techniques and the application of numerical models.
  
  - Examples
    - Model output from running a numerical climate model
    - Time series logged by environmental instrumentation
    - Conductivity-Temperature-Depth casts from oceanographic cruises
    - Groundwater chemistry and stable isotope measurements
    - Butterfly abundance observations

# ESRC data policy

---

- The [ESRC Research data policy](#) aligns with the RCUK Common Principles on Data Sharing
  - We expect grant holders to generate scientifically robust data ready for further re-use.
  - It is the grant holder's responsibility to incorporate data management as an integral part of the research project.
  - A structured approach to data management results in better quality data that is ready to deposit for further sharing

# The National Geoscience Data Centre (NGDC)



■ NERC Environmental Data Centre for Earth Sciences



# NGDC provides

- 
- Secure and appropriate storage for data and materials
  - Documentation, guides and advice
  - Support
  - Maintenance and management of confidentiality and release of data
  - Monitoring all correspondence and data delivery through the NERC's ishare DataMAD system
  - Data Citation

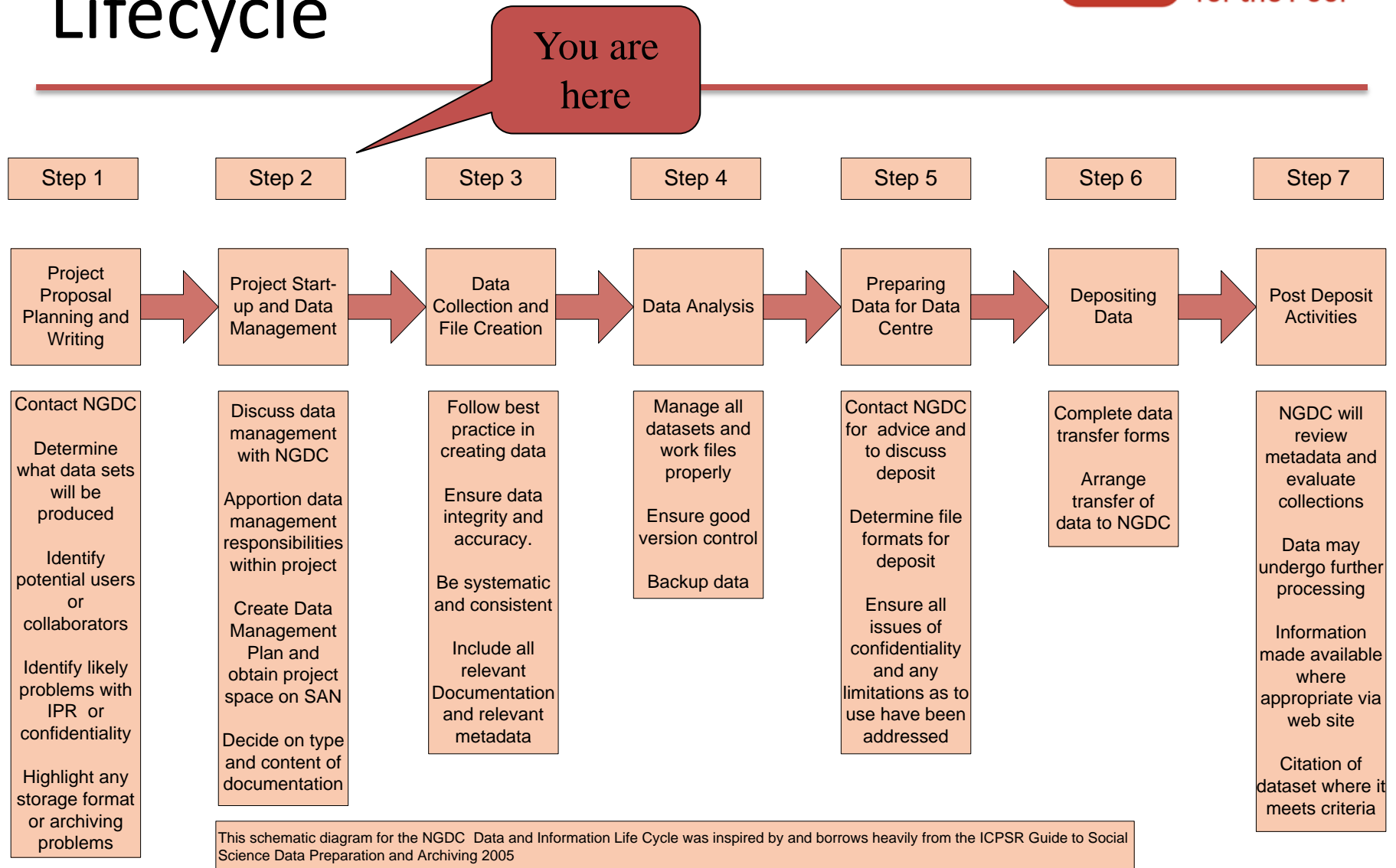


# Support

---

- We will provide guidance and help with DM issues (but won't do the data management within the projects)
- Advice about IPR and third party ownership
- Help with creating Data Management Plans for all parts of a project see: [Data Management Planning Guidance](#)
- Help with creation of metadata and their QA (both both for discovery and usage)
- Provide guidance on preferred data formats and ingestion mechanisms

# NGDC Data and Information Lifecycle



# What do we need to do?

- Data management activities need to be carried out throughout the lifetime of the project
- Data archiving is not only an 'end of project' activity
- Collection of metadata (information about your data sets) is collected by you
- Preparation of metadata is mostly easily achieved during the project
- Identify with you data of long-term interest as it is created

## Earth Science Academic Archive

Guidelines for  
preparing, donating  
and depositing data



NATURAL  
ENVIRONMENT  
RESEARCH COUNCIL

The National Geoscience  
Data Centre



British Geological Survey, Keyworth, Nottingham, NG12 5GG

# Check list

- Have you considered data backup and security
- Have you considered data format(s)
- Have you discussed all types of project output with the Data Centre
- Is Data Management a fully integrated part of the project management
- Will you supply data to the Data Centre before the end of the project for security



# Check list

---

- Have you completed a Data Management Plan (DMP)
- Has the DMP been approved by the Data Centre
- Have you identified Data Managers for each area within the project
- Has third party IPR and Copyright been properly addressed by the Programme and its partners
- Are there any confidentiality issues that need to be addressed
- Are you collecting sufficient metadata to ensure the data is of future value.

# Receiving and archiving deposited data

---

- Data should be delivered as per the Data Management plan
- We will be already be aware of any issues e.g.
  - What long term concerns are there with storage and preservation
- We will check whether the descriptive and technical metadata is good enough to describe this dataset uniquely?

# Data quality checks

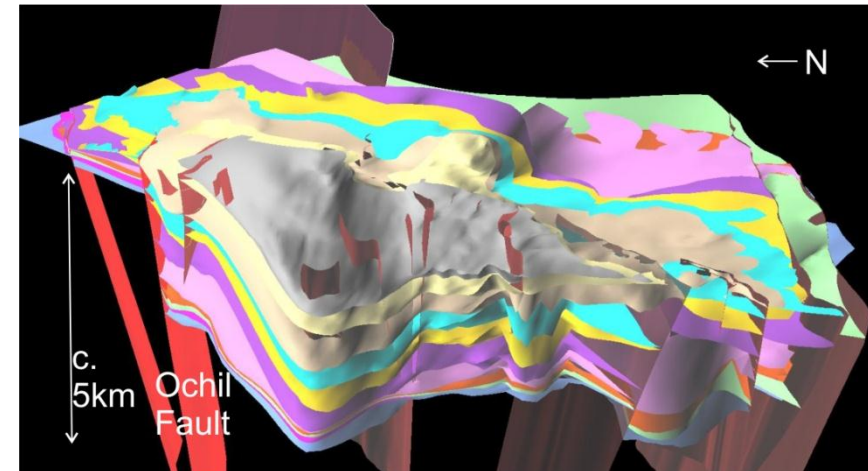
---

- The NGDC deals with a wide variety of datasets
- We may have no specific experience of your data
- We have to accept a lot on faith and work with you
- Many projects will create multiple files in multiple formats
- We will ensure data is conforms with the requisite formats, policies and any legal requirements

# Data formats/types

5/1/01  
Bram Moor Borehole

Sample	Bed / Member	Depth M	K+C	M	SM CP	Por. Smear 1st 8g core	Grain size 16-500 µm	Q <sub>2</sub>	K+C	M	SM <sub>1</sub>
BM 42	Concretion Beds	82.00	3	71	26	—	—	—	—	—	—
43	—	83.50	4	77	17	—	—	—	—	—	—
40	—	85.00	11	57	32	4.6	—	—	—	—	—
41	—	86.50	16	39	45	7.63	—	—	—	—	—
39	—	88.00	17	44	37	6.1	—	—	—	—	—
38	—	89.00	18	49	32	6.15	—	—	—	—	—
37	—	91.00	24	56	20	6.35	—	—	—	—	—
36	—	95.59	31	45	23	5.3	—	—	—	—	—
35	—	98.30	60	40	0	6.2	—	—	—	—	—
34	—	100.83	76	15	9	—	—	—	—	—	—
33	—	102.44	60	40	0	—	—	—	—	—	—
32	—	104.83	41	40	0	—	—	—	—	—	—
31	—	105.89	21	44	24	6.15	—	—	—	—	—
30	—	107.92	—	—	—	6.35	—	—	—	—	—
29	—	108.91	52	18	21	6.3	—	—	—	—	—
28	—	110.69	60	33	7	—	—	—	—	—	—
27	—	111.82	50	23	27	—	—	—	—	—	—
26	—	114.82	68	18	14	—	—	—	—	—	—
25	—	114.80	—	—	—	—	—	—	—	—	—
24	—	115.82	—	—	—	6.3	—	—	—	—	—
23	—	117.70	61	27	11	6.28	—	—	—	—	—
22	—	119.40	55	34	11	6.0	—	—	—	—	—
21	—	121.25	53	33	15	6.2	—	—	—	—	—
20	—	122.07	39	37	23	6.0	—	—	—	—	—
19	—	125.41	34	41	25	5.88	—	—	—	—	—
18	—	127.51	27	43	20	7.3	—	—	—	—	—
17	—	129.20	60	33	7	—	—	—	—	—	—
16	—	131.03	22	46	32	5.0	—	—	—	—	—
15	—	132.97	54	25	21	5.0	—	—	—	—	—
14	—	134.84	—	—	—	—	—	—	—	—	—
13	—	136.31	—	—	—	6.2	—	—	—	—	—



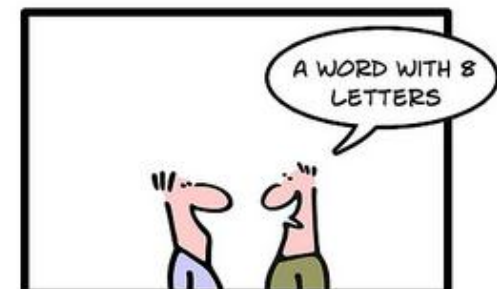
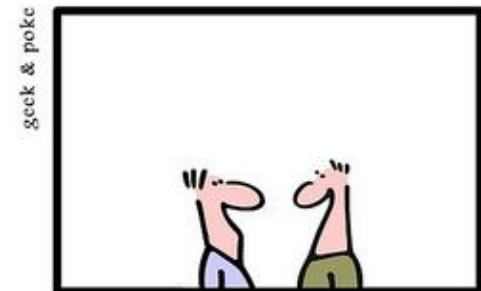
Name	Date modified	Type	Size
Electron Microprobe Data-Entrada sandstone	15/11/2012 15:11	Adobe Acrobat D...	73 KB
Green River Bulk rock chemistry data 1	21/05/2012 17:14	Microsoft Office E...	10,321 KB
internal rsd's trace element data1	21/05/2012 17:15	Microsoft Office E...	163 KB
Sequential leaching Entrada sandstone	15/11/2012 15:21	Microsoft Office E...	274 KB
Sequential leaching method for file - sequential leaching Entrada sandstone	15/11/2012 15:21	Adobe Acrobat D...	32 KB



# Metadata

- To ensure that research data can be effectively re-used by others, sufficient metadata must be recorded and published openly to ensure that research data are both discoverable and can be independently understood without recourse to the creator.

*SIMPLY EXPLAINED:  
METADATA*



geek & poke

# What is the Data Deposit and Search Portal

- <http://www.bgs.ac.uk/services/ngdc/esaa/>
- Set up to manage, within the NGDC, the data received from NERC Earth Science grants
- Expanded to include all similar data archived for academic research projects
- Discovery metadata to agreed International Standard
- Currently 250 projects on-line
- Confidentiality of data within agreed embargo period
- Knowledge of the availability of data helps potential collaboration and re-use

# Data Portal - Search



Unlocking the  
Potential of  
Groundwater  
for the Poor

The screenshot shows a web browser window displaying the British Geological Survey (BGS) website. The address bar shows the URL: <http://www.bgs.ac.uk/services/ngdc/esaa/downloads.html>. The page title is "Earth Science Academic Archive downloads". The main content area is titled "About the ESAA Downloads" and describes the archive as part of the National Geoscience Data Centre (NGDC). It provides information on how to search for projects and datasets, including text search and alphabetical lists. A sidebar on the left contains links to various data services, including "Our data", "Our products", "OpenGeoscience", "National Geological Repository", "National Geoscience Data Centre", "NGDC index", "Cited data", "Data management", "Digital data", "Earth Science Academic Archive", "Materials collection", "National Hydrocarbons Data Archive", "Records and archives", and "Premium data services". A right sidebar lists "Downloads" and "Search ESAA" options. The "Downloads" list includes: "Guidelines for depositors 528 KB pdf", "Guidelines for preparing, donating and depositing data 1.1MB pdf", "Data management questionnaire 100KB word .doc", "Data, materials and documentation transfer form 120KB word .doc", "Data management plan 180KB pdf", "The principles of good data management 550KB pdf", and "Quick check list 95KB word .doc". The "Search ESAA" section includes links to "Search ESAA data holdings" and "Terms of use". The "See also" section lists various resources like "Depositing data", "NGDC contacts", "Relocating UKCS Core", "Enquiries", "BGS downloads", "Rock Classification Scheme", "Vocabularies", and "NERC Open Research Archive". The "NERC data" section lists "NERC data centres".

mobile site

British Geological Survey  
NATURAL ENVIRONMENT RESEARCH COUNCIL

About us | Contact us | Downloads | Jobs | Shop

Search

Gateway to the Earth

Home | Our data | Our research | Our services | Our people | Discovering geology | News & Events | Hosted sites

Home » Our data » NGDC information & data » Earth Science Academic Archive » ESAA Downloads

## Our data

### Our products

OpenGeoscience

National Geological Repository

National Geoscience Data Centre

NGDC index

Cited data

Data management

Digital data

Earth Science Academic Archive

Materials collection

National Hydrocarbons Data Archive

Records and archives

Premium data services

## Earth Science Academic Archive downloads

### About the ESAA Downloads

The Earth Science Academic Archive has been set up as part of the National Geoscience Data Centre (NGDC) and is the place for you to deposit the results of your Earth Science research.

Since many projects have more than one grant award, to avoid duplication the data are stored within one project area. Details are given in the project abstracts.

### Search projects and datasets

Text search

Search project and dataset titles and abstracts

### Browse alphabetical lists

- A-Z list of principal investigators
- A-Z list of projects
- A-Z list of keywords

### Notes

#### Terms of use

All supplied documents and data files are available for direct download free of charge from the individual project pages within the ESAA data holdings.

*Terms of use* for data and information deposited in the Earth Science Academic Archive.

#### Help searching ESAA

In most cases the data available are exactly as deposited by researchers, but in some cases, to make the information readily available, the format may have been changed or the datasets grouped together.

Details of the projects and data are based on a hierarchical system so there will always be a project level abstract (even if no data is held in the ESAA). Individual subprojects or datasets within the project will also have an abstract description.

A search of the ESAA data holdings will return both projects and datasets with the datasets highlighted in blue. Selecting a particular dataset will allow you to view the project details and all of the associated individual datasets.

Where no data is held in the ESAA data holdings this will be stated in the Download details. Alternatively you may

### Downloads

- Guidelines for depositors 528 KB pdf
- Guidelines for preparing, donating and depositing data 1.1MB pdf
- Data management questionnaire 100KB word .doc
- Data, materials and documentation transfer form 120KB word .doc
- Data management plan 180KB pdf
- The principles of good data management 550KB pdf
- Quick check list 95KB word .doc

### Search ESAA

- Search ESAA data holdings
- Terms of use

### See also

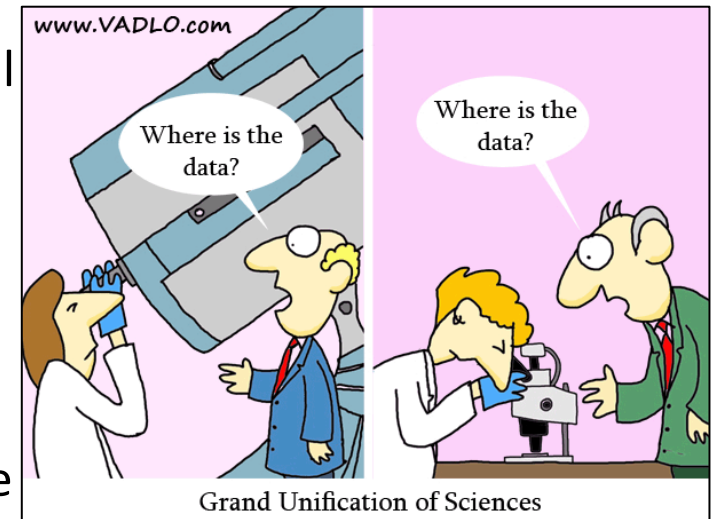
- Depositing data
- NGDC contacts
- Relocating UKCS Core
- Enquiries
- BGS downloads
- Rock Classification Scheme
- Vocabularies
- NERC Open Research Archive

### NERC data

- NERC data centres

# Data citation - why

- A DOI enables citation of datasets in the same manner as papers in a scientific journal
- It enables credit to be assigned to the dataset creators, and ensuring its discoverability and permanence
- Issuing DOIs and the publication of data recognises the value of the data and the effort that has gone into its creation, capture and effective management
- DOIs allow formal publication of the dataset in data journals.
- By assigning a DOI we're giving it our Data Centre stamp of approval, saying that it's of good quality, and that all the metadata is available



# Will all data get DOI

- 
- No – If it doesn't meet the requirements or a DOI is not requested.
  - The dataset will need to be permanent, by being deposited in NGDC.
  - The ownership of the dataset has to be clearly defined, along with the licensing terms to allow it to be made available.
  - The metadata describing the dataset on the 'landing page' must be complete and understandable.

# To sum up

---

- Take Data Management seriously – put as a standing item on project meeting agendas
- Complete and keep data management plans up to date
- Manage data properly and securely
- Make sure you deal with all confidentiality or third party issues
- Record appropriate and good metadata
- Deliver data to Data Centre according to the agreed schedule
- Don't hesitate to ask for help or support

# Contact and information

■ Rod Bowie – [rcb@bgs.ac.uk](mailto:rcb@bgs.ac.uk)

■ Information and Guidance:

- NERC Data Management Planning  
<http://www.nerc.ac.uk/research/sites/data/dmp/>
- NGDC – Earth Science Academic Archive  
<http://www.bgs.ac.uk/services/ngdc/esaa/>
- ESRC – Data Policy <http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx>
- NERC – Data Policy and guidance  
<http://www.nerc.ac.uk/research/sites/data/policy/>

■ Data Portal:

- <http://www.bgs.ac.uk/services/ngdc/esaa/downloads.html>