NGDC Data Value Checklist

Purpose and scope

The Data Value Checklist aims to identify which geoscientific or related data should be considered for accession to the National Geoscience Data Centre (NGDC) to derive the maximum value for science.

The data value checklist is not expected to give a definitive response to whether data should be retained, but will offer guidance on assessing its long-term value.

General guidance on the selection of data for long-term preservation

Selection of data should be based on quality, integrity, originality and geographic coverage. Data retained must tender a contribution to the scientific knowledgebase. The data may be used to inform national policy making or in an international context.

1. RELEVANCE TO MISSION
Is the data aligned with the NERC current strategy and falls in the environmental data remit of the NERC Data Centres? Consideration should be given to any legal or legislative requirements to retain data, for example, compliance with the Environmental Information Regulations and any contractual obligations which exist relating to the long term management and storage of data.

2. SCIENTIFIC OR HISTORICAL VALUE
Is there, or could there potentially in the future be, a use for this data? Could the material be scientifically or communally important? Is the data exemplary or does it set a precedent? Predicting future trends in research is difficult but consideration should be given to current trends in research awards and the scientific direction of research institutes alongside any educational value which might be obtained.

3. UNIQUENESS
Is this the primary and most complete unprocessed version of the data, to which no irreversible transformations have been applied? There are areas covered by NERC research where this approach may not be appropriate for example interpreted modelling, or 3D and 4D seismic data where the volumes of raw data are so large that post-stack or pre-stack are more suitable for long term retention and storage, however, these will be exceptions rather than the norm.

The NERC Data Centre will hold the principal copy of the data. Is the information new and unique or a re-working of previous material? If other copies existed would they be at risk or will they be preserved and if so they are the most complete and up to date?

4. NON-REPLICABILITY
Where it is not realistic to reproduce data, this is usually constrained by the costs incurred in the creation of the original dataset. Observations and sampling are often seen as non-repeatable, simulations could potentially be run again and experiments are repeatable, subject to cost implications.

5. POTENTIAL FOR REDISTRIBUTION
How reusable will the data be? Will it be stored in a format which will enable future re-use? Will the data be tied to a specific type of software? Will there be technical issues in reusing the data? Are there precautions which can be taken i.e. storing software alongside the data to future proof its reuse, licensing software models.

Are there any Intellectual Property Rights issues associated with the data and its reuse? Are there constraints would restrict future use of the data or are there any contractual or licensing terms which affect future use of the data? Is this an unaltered dataset which has not been changed in any way with its original integrity retained?

6. ECONOMIC CASE
When considering the preservation of data the cost of retention (identified not simply as storage but including managing, sharing, accessing, backing up and long term maintenance of data,) should be balanced against evidence of potential reuse of the data. A full economic case for retention will need to be made once a grant application is accepted. The Data Centre will need to consider the likely cost of preservation.

7. FULL DOCUMENTATION
Is there information i.e. completed metadata which will support the sharing, access and re-use of the data? In preparing this information considerations must be given to the fact that the individuals preparing the material for retention may not be involved in later projects which re-use the material.

Checklist

Essential criteria: These are legal or regulatory criteria and answering 'Yes' to **one or more** of the questions below will automatically result in selection for retention.

Legal/statutory considerations	**Yes**	**No**
Is there a legal or legislative reason for NERC to retain the data under any of the following:		
Science & Technology Act 1965		
Mining Industry Act (1926)		
Water Resources Act (1991)		
Petroleum Operations Notice 9 (PON 9) regulations (on-shore and off-shore)		
Public Records Act (1958 & 1967)		
Has or could the data been used in litigation, public enquiries, police investigations or any report or paper that could be legally challenged?		
Are there any financial or contractual obligations that require us to retain the data?		

Important criteria: These are primary criteria and answering 'Yes' to **at least one** of the questions from each section below should result in selection for retention.

Policy		
Does the NERC Data Policy apply to this data?		
Are the data a result of NERC/BGS funded activities?		
Does this data fall within the NGDC remit?		
Scientific or historic value		
Does the data have a geographical or temporal extent that makes it useful to others?		
Does the data have historic value i.e. does it represent a landmark in scientific discovery?		
Do the data include changes in processing methods, new standards or set any precedents?		
Do the data support current projects or trends in science?		
Is there likely to be further work in this or associated science areas?		
Are the data likely to meet the future needs/direction of the scientific community?		
Do the data contribute to a wider collection?		
Is there potential for re-use of the data?		
Are the data cited in a publication?		

Supporting criteria: These are important criteria and answering 'Yes' to **the majority** of the questions below should result in selection for retention.

Origin		
Are the data unique?		
Are the data unaltered with their original integrity retained?		
Would the data be costly to reproduce or reacquire?		
Is this thought to be the primary copy of this data?		
Are any copies of this data at risk?		
Condition		
Do the data have relevant metadata attached?		
Is there proportionally more data of scientific value in the collection than non-scientific value?		
Can the data be ingested into NGDC without additional processing? (reboxing, sifting, format conversion etc)		
Are the data in a suitable condition for addition to the collections? i.e. Readable, Undamaged, Uncorrupted, Robust enough to be handled		
Storage and preservation		
Can the data (digital or hardcopy) be stored without any exceptional requirements?		
Can the data (digital or hardcopy) be preserved without any exceptional requirements?		
Access/use		
Can the data be deposited without any restrictions in terms of intellectual property rights or copyright?		
Can the data be deposited without any established terms and conditions imposed by external sources?		
Can the data be deposited without any time limitations on its use?		
Formats/technical limitations		
Are the data in an acceptable format for deposit? (see table of accepted formats)		
Is the data accessible without any specialist (i.e. not commonly available) software?		
Is any specialist software readily available to the Data Centre?		
If not in acceptable format is it feasible for the data to be transferred to an appropriate storage/preservation or common use format?		