

## NERC Data Value Checklist

### Purpose and scope

The Data Value Checklist aims to identify which data should be considered for accession to the NERC Environmental Data Centres. The individual Data Centres' collections policies (both written and informal) will assist in deciding which Data Centre is the most appropriate place to deposit the data depending upon the science area and type of data collected.

The Data Value Checklist is intended to be used in the following circumstances: -

- a) When preparing a full Data Management Plan to assist Data Centres and Principal Investigators in determining the likely long term value of the data to be produced by a project.
- b) Upon receipt of the data for deposit with the Data Centres, to assess their quality, integrity, originality and content

This will ensure that data included in the NERC Data Centre collections are of long term value to the scientific community.

The Data Value Checklist is not expected to give a definitive response to whether the data should be retained, but will offer guidance on assessing their long-term value.

### Collection policy statements

Some Data Centres have a written Collections Policy on their website. Those that do not have a written policy will be able to advise whether data falls within their remit.

### Checklist

**Mandatory criteria:** These are mandatory 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?		
Is there any obvious reason why the data may be 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 probably result in selection for retention.

Policy		
Are the data a result of full or partial NERC funded activities?		
Do the data fall within the selected Data Centre's Collection Policy? If no – refer to NERC Data Coordinator or pass to the correct data centre.		
Scientific or historic value		
Are the data a unique unrepeatable measurement of the environment?		
Do the data have a broad geographical or temporal extent that makes them useful to others?		
Do the data have historic value i.e. do they 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?		
Are the data likely to meet the future needs/direction of the scientific community?		
Do the data contribute to a pre-existing collection?		
Is there potential for re-use of the data?		
Are the data likely to be cited or referenced 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		
Do the data have their original integrity?		
Would the data be costly to reproduce?		
Will this become the reference copy of the data?		
Condition		
Do the data have relevant metadata available?		
Are there proportionally more valuable data than non-valuable data within the collection?		
Can the data be ingested into the Data Centre without significant additional processing? (reboxing, sifting, conversion etc)		
Are the data in a suitable condition for addition to the collections? i.e. Readable, Undamaged,		

Uncorrupted, Robust enough to be handled		
<b>Storage and preservation</b>		
Can the data (digital or hardcopy) be stored without any exceptional requirements?		
Can the data (digital or hardcopy) be preserved without any exceptional requirements?		
<b>Access/use</b>		
Are the data free from any established terms and conditions imposed by external sources?		
<b>Formats/technical limitations</b>		
Are the data in an acceptable format for deposit?		
Are the data accessible without any specialist software?		
If Yes to the question above, Is any specialist software readily available to the Data Centre?		
Is it feasible for the data to be transformed to an appropriate format?		

## **Appendix 1 General guidance on the selection of data for long-term preservation**

Selection of data should be based on 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**

Are the data aligned with the NERC current strategy and fall 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 the data? Could the material be scientifically or communally important? Are the data exemplary or do they 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 with 3D and 4D seismic data 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, are they 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 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 they 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 their reuse.

Are there any Intellectual Property Rights issues associated with the data and their reuse? Are there constraints from the original funder which 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.