

DATA SELECTION AND APPRAISAL CHECKLIST

University of Reading Research Data Archive

Introduction

This Selection and Appraisal Checklist provides a set of criteria against which data offered or submitted to the University of Reading Research Data Archive will be assessed. The selection and appraisal criteria are specified in accordance with the Archive [Collection Policy](#).

Data offered or submitted to the Archive will be assessed for both eligibility and value. Data must be within the scope of the Collection Policy and of sufficient value to merit inclusion in the Archive. Data judged to be ineligible, trivial, of poor quality or without practical re-use value will not be accepted in the Archive. Where data are accepted for inclusion in the Archive, an appraisal of their value can help to inform decisions about long-term retention and when to review data for possible disposal.

A Dataset submitted to the Archive will also be assessed against the minimum acceptance criteria for a Dataset specified in the [Submission Policy](#), and may be rejected if it fails to meet these criteria.

Anyone wishing to deposit data in the Archive can use this Checklist to assess their own data. For guidance on making an assessment of data and further information about depositing data in the Archive, contact researchdata@reading.ac.uk.

Selection Criteria: Mandatory

These are mandatory criteria, and answering **Yes** to at least one question in each of sections 1, 2 and 3 below will result in the data being assessed for inclusion in the Archive.

ID	Criterion	Y / N /N/A	Notes/comments
Affiliation			
1	Have the data been collected or generated by at least one person who was a member of University staff or a registered postgraduate student at the University during the period of data collection?		
Eligibility			
2.1	Are the data necessary for the validation of published research findings/replication of published results, or are they likely to be necessary for this purpose?		
2.2	Do the data need to be retained to comply with any funder policy or the University Research Data Management Policy?		
2.3	Do the data need to be retained for legal or regulatory compliance reasons, e.g. to comply with Freedom of Information or Environmental Information regulations, or as part of a health and safety record?		
2.4	Do the data need to be retained to comply with any contractual terms and conditions, e.g. as specified by a commercial partner?		
2.5	Do the data have long-term value independent of any actual or intended research publication that may rely on or reference the data, for example, to support research re-use, teaching, commercial application, decision-making or policy formulation? (See Appraisal Criteria, below.)		

Data Selection and Appraisal Checklist

ID	Criterion	Y / N /N/A	Notes/comments
Ownership, rights and permissions			
3.1	Does the prospective depositor own the data or have permission to distribute the data on behalf of all rights-holders?		
3.2	If the prospective depositor needs to seek permission to distribute the data, can it reasonably be expected that this will be granted by the other rights-holders?		

Selection Criteria: Conditional

These criteria may not apply in all cases, but where they do, answering **Yes** to the questions in sections 4 and 5 will result in the data being assessed for inclusion in the Archive. Where a criterion is not applicable, this will not affect assessment for inclusion in the Archive either way.

ID	Criterion	Y / N /N/A	Notes/comments
Other services			
4	If a relevant, trusted, accessible domain- or data type-specific external data service exists, have the data been offered to this service and rejected? If so, the reasons for rejection should be supplied for purposes of appraisal.		
Consent and anonymization			
5.1	If the data were obtained from research participants, has consent for data sharing been obtained?		
5.2	If the data need to be anonymised for sharing, have they been or can they be effectively anonymised?		

Appraisal Criteria

These are important criteria for appraising the value of data that are eligible for inclusion in the Archive. Answering **Yes** to one or more of these should result in data being accepted. Where data have been rejected by a suitable external service, the reasons for rejection should be taken into account in the appraisal.

ID	Criterion	Y / N /N/A	Notes/comments
Uniqueness			
6	Are the data unique and unrepeatable, e.g. observational data?		
Repeatability			
7	If the data are repeatable, e.g. experimental or simulation data, would they be difficult and/or costly to reproduce or regenerate if they were lost?		
Simulations			
8	If the data are outputs of simulations, do they meet the archiving threshold specified in the BADC Model Data Policy? ¹ (The Policy concerns environmental simulation data, but its principles can apply to any analogue of real-world systems or processes generated by means of a computational model.)		
Scientific/historical value			
9	Do the data have evident scientific or historical value, i.e. do they represent a landmark in scientific discovery, do they include or demonstrate new processing methods or standards, do they establish any significant precedents, or support current or future research trends?		
Complementary/added value			
10	Do the data add value to or fill a gap in a pre-existing data collection or series? (E.g. a simulation may integrate observational data in a way that adds value to the observations.)		

¹ <https://badc.nerc.ac.uk/data/policies.html>.

ID	Criterion	Y / N /N/A	Notes/comments
Re-use value			
11	Do the data have any known or potential uses for research, teaching, commercial application, decision-making or policy formulation?		
Substantiveness			
12	Are the data substantive, in terms of geographic or temporal extent, number and range of data points, sufficient to make them useful to others? Some data may be too limited in extent, incomplete or trivial to have any practical scientific value.		
Domain standards			
13	Do the data conform to domain norms or requirements for the validation or replication of research results and meet accepted data standards or use conventional formats, where relevant? In some domains and some kinds of research, description of the research method and provision of the means to replicate results, e.g. software code/systems specification, may be more relevant and useful than archiving of outputs.		
Quality/usability			
14	Are the data of re-usable quality? Are they or can they be stored in identified and accessible formats that enable effective re-use through extraction and analytical processing? Are the data well-formed and appropriately documented? Are the provenance and transformations of the data clearly established?		

ID	Criterion	Y / N /N/A	Notes/comments
Access			
15	Can the data be made openly available? If terms and conditions imposed by any rights-holders mean data can only be made available on restricted terms, e.g. for non-commercial use or to authorised users only, this may limit the use-value of the data and affect compliance with open data policies of funders and publishers. Data may be embargoed for a limited period, e.g. pending publication or IP protection.		
Volume			
16	Is the volume of the dataset appropriate? All datasets greater than 10 GB in size must be approved prior to deposit in the Archive. The data should not include redundant duplication (e.g. where raw data have been processed through several stages), or unnecessarily large files. As a general rule, no individual file should be larger than 4 GB.		
Cost-effectiveness			
	Is it cost-effective to store the data in the Archive? Where data are high in volume, or where simulations are produced by a trivial algorithm that could be easily regenerated from a published description, the value of the data should be assessed against the cost of ingest and long-term storage.		