



# UK Data Service - Data management costing tool and checklist

The UK Data Service has prepared this costing tool and checklist to help formulate research data management costs in advance of research starting, for example for inclusion in a data management plan or in preparation for a funding application.

This tool considers the additional costs - above standard planned research procedures and practices - that are needed to preserve research data and make them shareable beyond the primary research team. The checklist indicates the activities to consider and cost to enable good data management. Such additional activities may require extra researcher or administrative staff time input, equipment, software, infrastructure or tools.

There are no hard and fast rules for costing data sharing requirements, as some research projects will pay more attention to detailed data documentation, organisation and formatting than others as part of routine fieldwork or preparation before analysis. Much also depends on the long-term storage, preservation and publication plans beyond the duration of the research itself. When data are deposited with a professional data centre or repository, such as the UK Data Archive, data preservation and dissemination activities are covered by the data centre/repository.

## How to use this costing tool

### Step 1:

Check the data management activities in the table and tick those that may apply to your proposed research.

### Step 2:

For each selected activity, estimate the additional time and/or other resources needed and cost this, e.g. people's time or physical resources needed such as hardware or software. Find out which resources, e.g. for data storage and backup, are available to you from your institution. Consider whether you need a dedicated data manager.

### Step 3:

Add these data management costs to your research application. Coordinate resourcing and costing with your institution, research office and institutional IT services.

### Step 4:

Plan the data management activities in advance to avoid them competing with the need to focus on research excellence.

### Caution:

Remember that when your research project nears the end you do not want these additional data management activities to compete with delivery of your planned outputs, writing of publications and the timely delivery of your project. At this later stage the costs of preparing data for sharing may be significantly higher.

## Reuse of this tool

We encourage sharing and reuse of these materials under the terms of the Creative Commons licence below. To cite:

UK Data Service (2013). Data management costing tool. UK Data Archive, University of Essex.



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported Licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

ACTIVITY	COMMENTS AND SUGGESTIONS	✓	COST
<p><b>Data description</b></p> <ul style="list-style-type: none"> <li>Are data in a spreadsheet or database clearly marked with variable and value labels, code descriptions, missing value descriptions, etc.?</li> <li>Are labels consistent?</li> <li>Do textual data like interview transcripts need description of context, e.g. included as a heading page?</li> </ul>	<ul style="list-style-type: none"> <li>if data description is carried out as part of data creation, data input or data transcription – low or no additional cost</li> <li>if needed to be added afterwards – higher cost</li> <li>codebooks for datasets can often be easily exported from software packages</li> </ul>		
<p><b>Data cleaning</b></p> <ul style="list-style-type: none"> <li>Do quantitative data need to be cleaned, checked or verified before sharing, e.g. check validity of codes used, check for anomalous values?</li> <li>Will data match documentation, e.g. same number of variables, cases, records, files?</li> <li>Does textual information in data need to be spell-checked?</li> </ul>	<ul style="list-style-type: none"> <li>if carried out as part of data entry and preparation before data analysis – low or no additional cost</li> <li>if needed afterwards – higher cost</li> </ul>		
<p><b>Documentation</b></p> <ul style="list-style-type: none"> <li>Do you have documentation for the data that describes the context and methodology of how data were gathered, created, processed and quality controlled?</li> </ul>	<ul style="list-style-type: none"> <li>often essential contextual and methods documentation will be written up in publications and reports</li> <li>if all data creation steps are well documented and documentation is kept well organised during research – low or no additional cost</li> <li>if documentation to be written or compiled specifically afterwards – higher cost</li> </ul>		
<p><b>Metadata</b></p> <ul style="list-style-type: none"> <li>Do structured metadata need to be created when data are shared via a data centre or archive, e.g. completing a deposit form for the UK Data Archive?</li> </ul>	<ul style="list-style-type: none"> <li>completing a UK Data Archive deposit form may take one to two hours</li> <li>other data centres will have their own metadata forms</li> </ul>		
<p><b>Formatting and organising</b></p> <ul style="list-style-type: none"> <li>Are your data files, spreadsheets, interview transcripts, records etc. all in a uniform format or style?</li> <li>Are files, records and items in the collection clearly named with unique file names and well organised?</li> </ul>	<ul style="list-style-type: none"> <li>if planned beforehand by developing templates and data entry forms for individual data files (transcripts, spreadsheets, databases) and by constructing clear file structures – low or no additional cost</li> <li>if needed afterwards – higher cost</li> <li>free software exists for batch file renaming to harmonise file names</li> </ul>		
<p><b>Transcription</b></p> <ul style="list-style-type: none"> <li>Will you transcribe qualitative data (e.g. recorded interviews or focus group sessions) as part of your research; or will you need to do this specifically so data can be more easily shared and reused?</li> <li>Is full or partial transcription needed?</li> <li>Is translation needed?</li> <li>Will you need to develop a standard transcription template or transcription guidelines, to ensure consistent formatting?</li> </ul>	<ul style="list-style-type: none"> <li>if part of research practice – very low or no additional cost</li> <li>if not planned as part of research practice – potentially high additional cost</li> <li>is additional hardware /software needed ?</li> <li>consider cost of (time needed for) developing procedures, templates and guidance for transcribers</li> <li>calculate time needed for transcription - four to eight hours per hour recording, e.g. using transcribing calculator: <a href="http://www.socialsciences.manchester.ac.uk/morgancentre/methods-and-resources/toolkits/toolkit-8">http://www.socialsciences.manchester.ac.uk/morgancentre/methods-and-resources/toolkits/toolkit-8</a></li> </ul>		



ACTIVITY	COMMENTS AND SUGGESTIONS	√	COST
<b>Digitisation</b> <ul style="list-style-type: none"> <li>Do analogue or paper-based research data (maps, newspaper clippings, photographs, images, text) need to be digitised to increase their potential for sharing?</li> </ul>	<ul style="list-style-type: none"> <li>is additional equipment or software needed for scanning or conversion?</li> <li>if simply image scanning of text – relatively low cost</li> <li>if Optical Character Recognition required, with manual checking for accuracy (revising entire scanned text) – may be high cost</li> <li>if manual data entry or typing needed, e.g. to digitise tabular data – may be high cost</li> </ul>		
<b>File format</b> <ul style="list-style-type: none"> <li>Do data need to be converted to a standard or open format with long-term validity for long-term preservation?</li> </ul>	<ul style="list-style-type: none"> <li>is additional software or hardware needed for conversion?</li> <li>for audio-visual data, converting to open digital formats can be time-consuming or require special equipment and/or software</li> <li>for databases, conversions may require checking for truncation, loss of metadata or annotation, loss of relationships, etc.</li> </ul>		
<b>Data storage</b> <ul style="list-style-type: none"> <li>How much data storage space is needed for the entire duration of the project?</li> <li>For long-term storage, decide which data will be kept long-term, which storage volume this represents and how long data will be stored and preserved.</li> </ul>	<ul style="list-style-type: none"> <li>if storage is provided by the institution – cost is included in standard indirect costs or overheads</li> <li>if additional storage needed – cost server/ disk space, as well as the cost of setting up and maintenance</li> </ul>		
<b>Data transfer and access</b> <ul style="list-style-type: none"> <li>Are special measures needed to transfer data from mobile devices, from fieldwork sites or from home equipment to a central work server?</li> <li>Do external people require access to research data?</li> </ul>	<ul style="list-style-type: none"> <li>is software or hardware needed for data transfer, for encryption of confidential data before transfer, or for synchronisation of data files across sites?</li> <li>does remote access via VPN or secure FTP need to be arranged for external people?</li> </ul>		
<b>Data backup</b> <ul style="list-style-type: none"> <li>Does the institution provide regular backup or not?</li> <li>Consider how frequently backups should be done, how many backups should be stored.</li> </ul>	<ul style="list-style-type: none"> <li>institutional backup – included in standard indirect cost /overheads</li> <li>additional backup needed – cost according to number of copies to be kept, frequency of backup and storage media needed</li> </ul>		
<b>Data security</b> <ul style="list-style-type: none"> <li>Protect data from unauthorised access or use or from disclosure</li> </ul>	<ul style="list-style-type: none"> <li>for confidential or sensitive data, determining conditions for controlling access to shared data may require extra time and discussion</li> <li>can security be arranged by institutional IT services or is extra software/hardware needed?</li> <li>data files may need encrypting before storage or transfers</li> </ul>		
<b>Consent for data sharing</b> <ul style="list-style-type: none"> <li>Do you need to ask participants for their consent for data to be shared?</li> <li>Consent is essential for qualitative interviews; less so in quantitative surveys where data can be more easily anonymised.</li> </ul>	<ul style="list-style-type: none"> <li>when consent for data sharing is considered as part of standard consent procedures early in research – very low or no additional cost</li> <li>when participants need to be re-contacted or re-visited after research has finished to obtain retro-active consent for data sharing – could be high cost</li> <li>does this require extra preparation of information sheets and consent forms; extra time for consent discussions; or training of interviewers?</li> </ul>		

ACTIVITY	COMMENTS AND SUGGESTIONS	√	COST
<p><b>Anonymisation</b></p> <ul style="list-style-type: none"> <li>Do you need to remove identifying information or conceal the identity of participants (e.g. using pseudonyms) before data can be shared?</li> <li>Anonymisation needs to be consistent throughout a data collection.</li> </ul>	<ul style="list-style-type: none"> <li>for quantitative data (e.g. survey data) – low cost if identifiers are a priori excluded from data files, are easy to remove, or identifiable variables are coded to avoid disclosure; cost may be higher if variables need recoding afterwards to avoid disclosure</li> <li>for qualitative textual data (e.g. interview transcripts) – may be high cost as entire texts will need to be read and checked for identifying information; costs can be reduced if anonymisation is carried out during transcription (or at least highlighted during transcription)</li> <li>for audio-visual data – anonymising/editing voices or faces can be very costly and reduces the usefulness of data</li> <li>cost depends on how sensitive or complex data are and how much identifying information is recorded in the data – if only removal of names is required, cost is low; pseudonymisation will require more time</li> <li>if anonymisation is planned before data collection or transcription/digitisation – cost can be lowered</li> <li>for all files, check file properties and edit to remove disclosive information such as editor/author name</li> </ul>		
<p><b>Copyright</b></p> <ul style="list-style-type: none"> <li>Do other parties hold copyright in the data?</li> <li>Do you need to seek copyright clearance before sharing data?</li> </ul>	<ul style="list-style-type: none"> <li>is time required to seek copyright clearance?</li> <li>is legal advice required?</li> </ul>		
<p><b>Data sharing</b></p> <ul style="list-style-type: none"> <li>Will your data be deposited with a data centre or institutional repository?</li> <li>Which requirements exist to prepare data to particular standards e.g. regarding documentation or format?</li> <li>Will journal publishers require deposit of data supporting article findings?</li> </ul>	<ul style="list-style-type: none"> <li>consider the cost of data deposit and/or longer-term storage – find out from data centre/repository/journal whether charges apply</li> <li>cost in time and effort needed to prepare the data for sharing and preservation</li> </ul>		
<p><b>Roles and responsibilities</b></p> <ul style="list-style-type: none"> <li>Do you need to allocate roles and responsibilities for various data management activities?</li> </ul>	<ul style="list-style-type: none"> <li>if multiple partner institutions, researchers or funders are involved in research – consider cost of data management planning meetings or discussions</li> </ul>		
<p><b>Operationalising data management</b></p> <ul style="list-style-type: none"> <li>What measures are needed to implement and operationalise data management throughout the research lifecycle?</li> </ul>	<ul style="list-style-type: none"> <li>do you need extra time and resources to implement data management throughout your research, e.g. regular team meetings, setting up a collaborative research environment?</li> <li>if staff training is required - higher cost</li> <li>do you need a dedicated data manager?</li> </ul>		