

## COSTING RESEARCH DATA MANAGEMENT IN THE SOCIAL SCIENCES

### ACTIVITY-BASED DATA MANAGEMENT COSTING TOOL FOR RESEARCHERS

This tool can be used to estimate the additional costs – above standard research procedures and practices – that are needed to make research data shareable beyond the primary research team. This can be done by listing all data management activities and steps required to make data shareable (e.g. based on a data management checklist), then costing each activity in terms of people's time or physical resources needed such as hardware or software. This model assumes that research data will be deposited with the UK Data Archive, which means that data storage, preservation and dissemination do not need to be included as the Archive covers those activities.

**Decide what are the additional costs, above your standard research procedures and practices, that would be needed to make your research data ready to be shared with other researchers.**

If you were to give your research data to another researcher who did not know your research, what additional edits, changes or upgrades would you need to make to the data, so this researcher can understand and use them?

The table indicates the activities you would list and cost as additional data management measures. The likelihood is that such additional activities require extra researcher or administrative staff time input.

1. Check the various suggested activities in the table and tick those that may apply to your research.
2. For each applicable activity, estimate the additional time and/or other resources needed and cost this.
3. Add these data management costs to your research application. Consider whether you need a dedicated data manager.
4. Remember that when your research project nears its end you do NOT want these additional data management activities to compete with your publications and the timely delivery of your project. At that stage the costs may be higher.
5. Plan the data management activities in advance to avoid them competing with the need to focus on research excellence.

| ACTIVITY   | ✓ | COMMENTS AND SUGGESTIONS   | COST |
|--|---|--|------|
| <p><b>CONSENT FOR DATA SHARING</b><br/><i>Do you need to ask participants for their consent for data to be shared? Essential for qualitative interviews; possibly less so in surveys where data can be more easily anonymised.</i></p> |   | <ul style="list-style-type: none"> <li>• When consent for data sharing is considered as part of standard consent procedures – very low or no additional cost</li> <li>• When participants need to be re-contacted or re-visited after research to obtain retro-active consent for data sharing – could be high cost</li> <li>• A possible cost to research may be difficulty to recruit research participants, if data sharing is discussed</li> <li>• In longitudinal research the cost may decrease over time</li> <li>• Does this require preparation of additional information sheets, consent forms or extra time for consent discussions/staff training?</li> </ul>  |      |
| <p><b>ANONYMISATION</b><br/><i>Do you need to remove identifying information or conceal the identity of participants before data can be shared? Anonymisation needs to be consistent throughout a data collection.</i></p>             |   | <ul style="list-style-type: none"> <li>• For quantitative data (e.g. survey data) – may be low cost</li> <li>• For qualitative textual data (e.g. interview transcripts) – may be high costs as entire texts will need to be read and checked</li> <li>• For audiovisual data – may be very high cost. Access controls may be preferred to prevent loss of data quality,</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.</li> <li>• If anonymisation is planned before data collection or data digitisation – cost can be lowered</li> </ul> |      |
| <p><b>DATA SECURITY AND ACCESS</b><br/><i>Protect data from unauthorised access, use, and disclosure</i></p>   |   | <ul style="list-style-type: none"> <li>• For sensitive qualitative data, deciding conditions for access control of shared data may require extra time, team discussion and discussion with the Archive.</li> </ul>   |      |
| <p><b>DIGITISATION</b><br/><i>Do analogue or paper-based research data need to be digitised to increase their potential for sharing?</i></p>   |   | <ul style="list-style-type: none"> <li>• If Optical Character Recognition and therefore checking for accuracy needed – may be high cost</li> <li>• If simply image scanning – may be low cost.</li> <li>• Is extra equipment or software needed?</li> </ul>  |      |
| <p><b>TRANSCRIPTION</b><br/><i>Do you transcribe qualitative data (e.g. interviews) as part of your analysis; or would you need to do this specifically so data can be shared and re-used?</i></p>                                     |   | <ul style="list-style-type: none"> <li>• If part of research practices – very low or no additional cost</li> <li>• If not planned as part of research practices – potentially high additional cost</li> <li>• Consider cost of developing procedures and templates and guidance for transcribers</li> <li>• Is full or partial transcription needed? Is translation needed?</li> <li>• If working with many transcribers or services, costs may be higher</li> </ul>   |      |

| ACTIVITY   | ✓ | COMMENTS AND SUGGESTIONS   | COST |
|--|---|--|------|
| <p><b>FORMATTING AND ORGANISING</b><br/><i>If your data collection consists of multiple files, are your data files, spreadsheets, transcripts, etc. all in a uniform format or style and named with unique and clear file names?</i></p> |   | <ul style="list-style-type: none"> <li>• If planned in advance by developing transcription template and guidelines or designing database format - low cost</li> <li>• If needed afterwards - higher cost</li> </ul>  |      |
| <p><b>LABELLING AND CODING</b><br/><i>Are quantitative data in a spreadsheet or database clearly labelled with variable and value labels and descriptions, code descriptions, missing value descriptions, etc.?</i></p>                  |   | <ul style="list-style-type: none"> <li>• If carried out as part of database design and development - low cost</li> <li>• If needed afterwards - higher cost</li> </ul>   |      |
| <p><b>CLEANING</b><br/><i>Do quantitative data need to be cleaned, checked or verified? Do qualitative data need to be spell-checked?</i></p>  |   | <ul style="list-style-type: none"> <li>• If carried out as part of data entry and preparation for analysis - low cost</li> <li>• If needed afterwards - higher cost</li> </ul>   |      |
| <p><b>DESCRIPTION</b><br/><i>Do textual data, e.g. interview transcripts, need context description?</i></p>  |   | <ul style="list-style-type: none"> <li>• If carried out as part of data creation, or data analysis - low cost</li> <li>• If needed afterwards - higher cost</li> </ul>   |      |
| <p><b>DOCUMENTATION</b><br/><i>Do you have documentation for the data that describes the methodology of how data were gathered, processed, quality controlled, etc?</i></p>  |   | <ul style="list-style-type: none"> <li>• If all data creation steps are well documented and documentation is kept well organised during research - low cost</li> <li>• If documentation needs to be written specifically afterwards - higher cost</li> </ul> |      |

| ACTIVITY  | ✓ | COMMENTS AND SUGGESTIONS   | COST |
|---|---|--|------|
| <b>METADATA</b><br><i>Do 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?</i>                         |   | <ul style="list-style-type: none"> <li>Completing a UK Data Archive data review form and deposit form may take 1-2 hours.</li> </ul>   |      |
| <b>FILE FORMAT</b><br><i>Do data need to be converted to a standard or open format with long-term validity?</i>   |   | <ul style="list-style-type: none"> <li>For audiovisual data, converting to open digital formats can be time-consuming or require special equipment or software</li> <li>Is additional software or hardware needed for conversion?</li> </ul> |      |
| <b>COPYRIGHT</b><br><i>Do other parties hold copyright in the data; if so, do you need to seek copyright clearance?</i>   |   | <ul style="list-style-type: none"> <li>Is time required to seek copyright clearance?</li> <li>Is legal input required?</li> </ul>  |      |
| <b>PLANNING, ROLES AND RESPONSIBILITIES</b><br><i>Do you need to plan data management for your research and allocate roles and responsibilities for various data management activities?</i> |   | <ul style="list-style-type: none"> <li>If multiple partner institutions, researchers or funders are involved in research - higher cost</li> </ul>  |      |
| <b>OPERATIONALISING</b><br><i>What measures are needed to implement and operationalise data management throughout the research cycle?</i>   |   | <ul style="list-style-type: none"> <li>Does putting your planned data management into practice throughout your research take time, e.g. for regular team meeting</li> </ul>  |      |

[www.data-archive.ac.uk/create-manage/projects/jisc-dmp](http://www.data-archive.ac.uk/create-manage/projects/jisc-dmp)