

Estimated costs apply to the Danish situation and need to be adjusted in other countries.

RDM TOOL FOR ESTIMATING COSTS AT SDU			
DMP phase	ACTIVITY	COMMENTS AND SUGGESTIONS	COSTS
1. Preparing	Make a Data Management Plan (DMP)	<ul style="list-style-type: none"> Make a DMP before you start creating data; make decisions about managing your data; consider how you can process, analyse, preserve and share your data . A nice tool is DMPonline. Seek support for data management planning at the SDU RDM-support page. 	<p>Small project: 2-4 hours .</p> <p>Big project: 2 days or more, depending on the complexity of your project.</p>
2.1 Data Collection	<p>Acquiring external datasets</p> <p>Do you plan to use existing data, and is the data available at a commercial partner?</p>	<ul style="list-style-type: none"> SDUB can help you acquire a license to a crucial database. See here the databases we have subscription to. In research data repositories, data can be available at no or low costs. Examples: https://datacite.org/, https://zenodo.org/, https://www.re3data.org/ . 	<p>Example: A faculty license on a database for macro-economic analyses: €18.000/y.</p>
2.2 Data Collection	<p>Formatting and organising</p> <p>Are your data files, spreadsheets, measurements, interview transcripts, records etc. all in a uniform format or style?</p> <p>Are files, records and items in the collection clearly named with unique file names and well organised?</p>	<ul style="list-style-type: none"> 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 . If needed afterwards –higher cost. 	<p>Per project organize style, format, names can be done by a student assistant at level 1* salary or research data manager at level 2* salary.</p>
2.3 Data Collection	<p>Transcription</p> <p>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</p>	<ul style="list-style-type: none"> If part of research practice –very low or no additional cost. If not planned as part of research practice –potentially high additional cost. Is additional hardware /software needed ? 	<p>Example: Time needed for transcription -four to eight hours per hour recording, i.e. see https://www.danish-transcribers.dk/.</p> <p>Or it could be done by a student</p>

	<p>can be more easily shared and reused? Is full or partial transcription needed?</p> <p>Is translation needed?</p> <p>Will you need to develop a standard transcription template or transcription guidelines, to ensure consistent formatting?</p>	<ul style="list-style-type: none"> Consider cost of (time needed for) developing procedures, templates and guidance for transcribers. 	<p>assistant at level 1* salary.</p> <p>Otherwise, you can use the paid SDU service of automatic transcription KONCH costing DKK 2075 DKK per year. Contact Louise Haugsted Knudsen.</p>
2.4 Data Collection	<p>Consent for data sharing</p> <p>Do you need to ask participants for their consent for data to be shared?</p> <p>Consent is essential for research in the domain of health/life sciences also for qualitative interviews</p>	<ul style="list-style-type: none"> When consent for data sharing is considered as part of standard consent procedures early in research – very low or no additional cost. When participants need to be re-contacted or re-visited to obtain - active consent–could be high cost. Does this require extra preparation of information sheets and consent forms; extra time for consent discussions; or training of interviewers? 	<p>Student assistant at level 1* salary or research data manager at level 2* salary.</p>
2.5 Data Collection	<p>Data transfer</p> <p>Are special measures needed to transfer data from mobile devices, from fieldwork sites or from home equipment to a central work server?</p>	<ul style="list-style-type: none"> Is software or hardware needed for data transfer, for encryption of confidential data before transfer, or for synchronisation of data files across sites? 	<p>SDU has an agreement with Nextcloud which can be used for free.</p>
3.1 Data Documentation	<p>Data description and metadata</p> <p>Are data in a spreadsheet, database or data warehouse clearly marked with</p>	<ul style="list-style-type: none"> If data description is carried out as part of data creation, data input or data transcription –low or no additional cost. If needed to be added or harmonized afterwards –higher cost. 	<p>Examples: 4 hrs per single experiment (120 measurements) filling in 60 required metadata fields, with assistance of a</p>

	<p>variable, variable labels and value labels, code descriptions, missing value descriptions, etc.?</p> <p>Are validated questionnaires and standard coding used?</p> <p>Are labels consistent?</p> <p>Are files, records and items in the collection clearly described with well-defined metadata or a metadata standard to interpret the relations between them and to quickly select and understand the content.</p> <p>Do textual data like interview transcripts need description of context, e.g. included as a heading page?</p>	<ul style="list-style-type: none"> • Codebooks for datasets can often be easily exported from software packages (eg, guides for REDCAP, guides for R). • Is a specific vocabulary, taxonomy or ontology followed for data and/or metadata? 	<p>research data manager at level 2* salary.</p> <p>Two to three weeks are costed into an average two year research grant application to prepare and collate materials for deposit (http://www.data-archive.ac.uk/help/user).</p>
3.2 Data Documentation	<p>Documentation</p> <p>Do you have documentation for the data that describes the context and methodology of how data were gathered, created, processed and quality controlled?</p>	<ul style="list-style-type: none"> • Often essential contextual and methods documentation will be written up in publications and reports. • If all data creation steps are well documented and documentation is kept well organised during research – low or no additional cost • If documentation to be written or compiled specifically afterwards – higher cost. 	<p>Researcher at level 2* salary.</p>
4.1 Data Storage & Back-up	<p>Data backup</p>	<ul style="list-style-type: none"> • Institutional backup –included in standard indirect cost /overheads. 	<p>Examples: University drive €0.80 per GB/y</p>

	<p>Does the institution provide regular backup or not?</p> <p>Consider how frequently backups should be done, how many backups should be stored.</p>	<ul style="list-style-type: none"> • Additional backup needed –cost according to number of copies to be kept, frequency of backup and storage media needed. 	<p>Cloud: €0.30 per GB/y 2 x Harddrive: €0.14 per GB (single purchase).</p>
4.2 Data Storage & Back-up	<p>Data storage</p> <p>How much data storage space is needed for the entire duration of the project?</p> <p>Do you need to set up a data model and accompanying database for the data?</p>	<ul style="list-style-type: none"> • If storage is provided by the institution –cost is included in standard indirect costs or overheads. • If additional storage needed –cost server/ disk space, as well as the cost of setting up and maintenance. • Do you need a data warehouse or a database architect? 	<p>SDU offers free storage at OneDrive and with Ucloud you get a credit of 5000 DKK for storage and 3000 DKK for computing.</p>
5.1 Data Access & Security	<p>Data access</p> <p>Do external people require access to research data?</p>	<ul style="list-style-type: none"> • Does remote access via VPN or secure FTP need to be arranged for external people? • Via Sharepoint external collaborators may be able to access SDU projects, as well as via Ucloud as it has a WAYF login. 	<p>Mostly researchers can make use of existing, free services.</p>
5.2 Data Access & Security	<p>Data security</p> <p>Is there an institutional server available where you can store your data safely?</p> <p>Protect data from unauthorised access or use or from disclosure</p>	<ul style="list-style-type: none"> • For confidential or privacy sensitive data, determining conditions for controlling access to shared data may require extra time and discussion. • Can security be arranged by institutional IT services or is extra software/hardware needed? • Data files may need encrypting before storage or transfers. 	<p>SDU offers for free Onedrive and Sharepoint as well as Ucloud.</p>
6. Data Preservation & Archiving	<p>File format</p> <p>Do data need to be converted to a standard or open format with long-</p>	<ul style="list-style-type: none"> • Is additional software or hardware needed for conversion? • For audio-visual data, converting to open digital formats can be time- 	<p>Researcher at level 2* salary.</p>

	<p>term validity for long-term preservation?</p>	<p>consuming or require special equipment and/or software.</p> <ul style="list-style-type: none"> For databases, conversions may require checking for truncation, loss of metadata or annotation, loss of relationships, etc. Specific forms need to be filled for preservation at the National Archives regarding the documentation, and the description of the file. 	
<p>7.1 Data Sharing & Reuse</p>	<p>Anonymisation</p> <p>Do you need to remove identifying information or conceal the identity of participants (e.g. using pseudonyms) before data can be shared?</p> <p>Anonymisation needs to be consistent throughout a data collection.</p>	<ul style="list-style-type: none"> If anonymisation is planned before data collection or transcription/digitisation –cost can be lowered . For audio-visual data – anonymising/editing voices or faces can be very costly and could reduce the usefulness of data. 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. For qualitative textual data (e.g. interview transcripts) –costs can be reduced if anonymisation is carried out during transcription (or at least highlighted/coded during transcription). 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. For files received of participants, check file properties and edit to 	<p>Free software is available. AMNESIA is a data anonymization tool that allows to remove identifying information from data.</p> <p>Example: Transcribing / simultaneously anonymizing audio (speech): up until one hour per 5 minute fragment (depending on the preciseness level of transcribing).</p> <p>Student assistant at level 1* salary.</p>

		<p>remove disclosive information such as editor/author name.</p> <ul style="list-style-type: none"> • For more information on working with sensitive data see the guides of Institute of Public Health. • For guides on how to process personal data, see the Databeskyttelse på SDU. 	
7.2 Data Sharing & Reuse	<p>Copyright</p> <p>Do other parties hold copyright in the data?</p> <p>Do you need to seek copyright clearance before sharing data?</p>	<ul style="list-style-type: none"> • Is time required to seek copyright clearance? • Is legal advice required? 	<p>Juridical advice at level 3* salary, or SDU RIO and SDU RDM-support can give free advice.</p>
7.3 Data Sharing & Reuse	<p>Data sharing</p> <p>Will your data be deposited with a data centre or institutional repository?</p> <p>Which requirements exist to prepare data to particular standards e.g. regarding documentation or format?</p> <p>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?</p> <p>What data will be retained and what not?</p>	<ul style="list-style-type: none"> • According to the Danish Code of Conduct for Research Integrity, data should in general be kept for a period of at least five years from the date of publication. • A research data repository/ data centre/ journal can provide you with the possibility to share your data for reuse. Find out what the cost are of data deposit and/or longer-term storage per year cost in time and effort needed to prepare the data for sharing and preservation. • Data centres will have their own metadata forms. Consider using these on beforehand. 	<p>Examples:</p> <p>Completing a data repository upload form (i.e. 3TU Datacentrum or DANS, or Zenodo free-of-charge repository) may take 15 min to 4 hours.</p> <p>Dryad €110 once (max 20 GB).</p> <p>DataverseNL €3.60 per GB/year</p> <p>Cloud Database as a service: €160 /month (storage 5 GB, transfer 30 GB).</p> <p>In case of big datasets with lots of metadata, consider also the cost of a student assistant at level 1* salary to just type and fill in all the</p>

			information at the repository. Sometimes different datasets need to be submitted to different repositories depending on the type of article or the demands of the publisher/journal.
7.4 Data Sharing & Reuse	<p>Data cleaning</p> <p>Do quantitative data need to be cleaned, checked or verified before sharing, e.g. check validity of codes used, check for anomalous values?</p> <p>Will data match documentation, e.g. same number of variables, cases, records, files?</p> <p>Does textual information in data need to be spell-checked?</p> <p>Do you need to combine your data with other datasets for your research?</p>	<ul style="list-style-type: none"> • Data cleaning takes time. • If carried out as part of data entry and preparation before data analysis –low additional cost. • If needed afterwards –higher cost. 	<p>Example: Data cleaning service: €270 to well over €1800. More info on http://datascopic.net/cost-of-data-cleansing/</p> <p>There are tools that can help you clean your data like Open Refine.</p> <p>Researcher/ research data manager at level 2* salary.</p>
7.5 Data Sharing & Reuse	<p>Digitisation</p> <p>Do analogue or paper-based research data (maps, newspaper clippings, photographs, images, text) need to be digitised to increase their potential for sharing?</p>	<ul style="list-style-type: none"> • Is additional equipment or software needed for scanning or conversion? • If simply image scanning of text – relatively low cost. • If Optical Character Recognition required, with manual checking for accuracy (revising entire scanned text) –may be high cost. 	<p>Example: Digitisation €0.50 per page (few pages) <i>OR</i> €320-390 per 1000 pages (OCR included).</p> <p>The Center for Special Collections and Digital Humanities at SDUB can help you</p>

		<ul style="list-style-type: none"> If manual data entry or typing needed, e.g. to digitise tabular data – may be high cost. 	<p>with most of these tasks.</p> <p>Student assistant at level 1* salary.</p>
8.1 Overall	<p>Roles and responsibilities</p> <p>Do you need to allocate roles and responsibilities for various data management activities?</p>	<ul style="list-style-type: none"> If multiple partner institutions, researchers or funders are involved in research –consider cost of data management planning meetings or discussions . 	Travel costs, lunch, time .
8.2 Overall	<p>Operationalising data management</p> <p>What measures are needed to implement and operationalise data management throughout the research lifecycle?</p>	<ul style="list-style-type: none"> 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? If staff training is required -higher cost. Do you need a dedicated data manager? 	Research data manager at level 2* salary.

* Salary:

Level 1 (i.e. student assistant) ~17 euro per hour (120-180 DKK).

Level 2 (researcher, data manager) ~60 euro per hour.

Level 3 (external expert) ~160 euro per hour.

If this table seems too complicated, just remember that on average, **5%** of overall research costs should go towards data stewardship and towards ensuring that data are reusable (Barend Mods, 2020).

This guide was based on the work of:

- [1] UK Data Service (2013). Data management costing tool. UK Data Archive, University of Essex.
- [2] Alisa Westerhof (UU), Tessa Pronk (UU), Annemiek van der Kuil (3TU & TUD), Annemie Mordant (UM) (2015). Data Management Bij wetenschappelijk onderzoek méér dan alleen storage. Landelijk Coördinatiepunt Research Data Management, The Netherlands
- [3] How to identify and assess Research Data Management (RDM) costs
<https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>
- [4] B. Mons (2020), Invest 5% of research funds in ensuring data are reusable, Nature 578, 491, doi:
<https://doi.org/10.1038/d41586-020-00505-7>