

## **Research Data Management Costing Guide**

### **Introduction:**

Research data management is an increasingly critical element of successful research and is now frequently required by funders. This means that data management elements are increasingly allowable as direct costs in grant applications. However, identifying what aspects of data management may incur a cost can prove difficult, especially early in the research process.

### **How to use this guide:**

This guide is divided into seven phases of the research lifecycle, with some stages having multiple considerations. Each phase presents one or more potential Activities. These Activities will often include guiding questions that may be familiar if you have already completed a data management plan. The Activities are followed by one or more Comments or Suggestions. Though some Comments are phrased as questions, they are more commonly statements that help explain when or how data related costs may be increased or decreased.

### **Limitations:**

This guide should be considered a starting point to begin thinking and having conversations about data management costs. You will need to consult your specific request for proposal or directorate and speak with an Office of Sponsored Programs representative to determine what costs are allowable in your circumstance and through your funder. The salary rates suggested are very general estimates and may not reflect your situation. The Albertsons Library Research Data Management Group is happy to consult with you about any of these elements, but cannot provide financial advice.

DMP Phase	Activity	Comments and Suggestions	Costs
1. Preparing	Make a data management plan	<ul style="list-style-type: none"> <li>• Make a DMP <i>before</i> you start creating data; make decisions about managing your data; consider how you can process, analyze, preserve, and share your data</li> <li>• Ensure that your DMP and IRB, IACUC, or IBC documentation are in alignment</li> <li>• Always consult with any on-campus departments you reference (e.g. Library, Research Computing, etc.) before submission</li> <li>• Albertsons Library offers consultations for DMP development</li> </ul>	<p>(Time Costs) Highly variable - 2 hours to 2 days depending on experience and the complexity of the project</p> <p>(Preferred) Allow 3 business days for DMP consultations<sup>1</sup></p>
2.a. Data Collection	<p>Acquiring External Datasets</p> <ul style="list-style-type: none"> <li>• Do you plan to use existing data?</li> <li>• Is the data available from an open access repository?</li> <li>• Is the data available from a commercial vendor?</li> </ul>	<ul style="list-style-type: none"> <li>• Albertsons Library may be able to help you locate existing datasets that support your research</li> <li>• Even open access existing datasets may require IRB or IACUC approval prior to use</li> <li>• If using outside data, you must fill</li> </ul>	<p>Dataset licenses are highly variable from a few hundred dollars to tens of thousands<sup>3</sup></p> <p>Existing datasets may need <i>Formatting and Organizing</i> (See 2.b)</p>

<sup>1</sup> Per “Basics of Charging Costs to a Sponsored Project” Section IV.K. Proposal Costs (200.460), “The cost of preparing proposals or applications to sponsoring agencies is not allowable as a direct cost. This includes the development of data necessary to support the proposal.”

<https://www.boisestate.edu/research-osp/charging-costs/>

<sup>3</sup> Per “Basics of Charging Costs to a Sponsored Project” Section IV.F. Memberships and Subscriptions (200.454), “...subscriptions to business, technical, and professional periodicals are allowable when they provide a direct benefit to the project.” Though acquiring data necessary to support the proposal is not allowable, licensing existing datasets as part of the actual research project may be allowable.

	<ul style="list-style-type: none"> <li>• If using an existing dataset, are there specific requirements or conditions set by the dataset owner?</li> </ul>	<p>out the Research Data Classification Form<sup>2</sup> to ensure the university has the proper measures in place</p>	
2.b. Data Collection	<p>Formatting and Organizing</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 organized?</li> </ul>	<ul style="list-style-type: none"> <li>• If planned beforehand by developing lab manuals, read-me files, and data entry forms that include clear file structures - low or no additional cost</li> <li>• If needed after data has been collected - higher cost</li> </ul>	<p>Depending on the project this can be done by a Student Research Technician 1, 2, or 3 at a salary class of S2, S3, or S4 or a Graduate Assistant under a stipend<sup>4</sup></p> <p>Complex projects may require a Data Coordinator or Research Analyst<sup>5</sup></p>
2.c. Data Collection	<p>Transcription</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</li> </ul>	<ul style="list-style-type: none"> <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> </ul>	<p>For estimated time costs and other considerations:</p> <ul style="list-style-type: none"> <li>• Transcribing Your Own Qualitative data<sup>6</sup></li> <li>• Using an External Agency to Transcribe Your Qualitative Data<sup>7</sup></li> </ul>

<sup>2</sup> Office of Sponsored Programs. (2022). *Research Data Classification Form FAQs*. Retrieved from <https://www.boisestate.edu/research-osp/rdcf-faqs/>

<sup>4</sup> Graduate College. (2022). *Graduate Assistant Work Obligation*. Retrieved from <https://www.boisestate.edu/graduatecollege/funding/graduate-assistants/ga-work-obligation/> and Human Resource Services. (2022). *Student Employment Classification Schedule*. Retrieved from <https://www.boisestate.edu/hrs/student-employment-classification-schedule/>

<sup>5</sup> Idaho Division of Human Resources. (2022). *Class Specifications*. Retrieved from <https://www.governmentjobs.com/careers/idaho/classspecs>

<sup>6</sup> Burke, H., L. Jenkins, and V. Higham. (2010). *Toolkit 08: Transcribing your own qualitative data*. Retrieved from <https://www.socialsciences.manchester.ac.uk/morgan-centre/research/resources/toolkits/toolkit-08/>

<sup>7</sup> Burke, H. (2011). *Toolkit 15: Using an external agency or individual to transcribe your qualitative data*. Retrieved from <https://www.socialsciences.manchester.ac.uk/morgan-centre/research/resources/toolkits/toolkit-15/>

	<p>more easily shared and reused?</p> <ul style="list-style-type: none"> <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>		
2.d. Data Collection	<p>Consent for Data Sharing</p> <ul style="list-style-type: none"> <li>• Do you need to ask participants for their consent for data to be shared?</li> <li>• If you intend to share or reuse data that requires informed consent documentation, you must outline these intentions in the consent documentation</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 - low or no additional cost</li> <li>• When participants need to be re-contacted or re-visited to obtain active consent - potentially high additional 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>	<p>If considered part of consent procedures throughout the project - incorporated into the cost of your data collection staff</p> <p>If participants must be re-contacted or re-visited this can usually be done by a Student Research Technician 2 or 3 with CITI Training or a Graduate Assistant with CITI Training under a stipend</p>
2.e. Data Collection	<p>Data Transfer</p> <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> </ul>	<ul style="list-style-type: none"> <li>• Is software or hardware needed for the data transfer, for encryption of confidential data before transfer, or for synchronization of data files across sites?</li> </ul>	<p>Research Computing can provide file transfer assistance including Globus access</p>
3.a. Data	Data Description and Metadata	<ul style="list-style-type: none"> <li>• If data description is carried out as</li> </ul>	The UK Data Archive

Documentation	<ul style="list-style-type: none"> <li>• Are data in a spreadsheet, database, or data warehouse clearly marked with variables, variable labels, and value labels, code descriptions, missing value descriptions, etc.</li> <li>• Are validated questionnaires and standard coding used?</li> <li>• Are labels consistent?</li> <li>• 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 content?</li> <li>• Do textual data like interview transcripts need description of context, e.g. included as a heading page?</li> </ul>	<p>part of data creation - low or no additional cost</p> <ul style="list-style-type: none"> <li>• If needed to be added or harmonized afterwards - higher cost</li> <li>• Codebooks for datasets can often be easily exported for software packages</li> </ul>	<p>recommends allowing for, on average, two to three weeks for a two-year grant application to prepare and collate materials for deposit into an archive or repository.<sup>8</sup></p> <p>The staff cost associated will depend on who has the necessary knowledge to complete the descriptions. If codebooks and labels are consistent a Graduate Assistant under stipend or Student Research Technician 2 or 3 at a salary rate of S3 or S4 can complete this task.</p> <p>If the researcher or research team are the only staff who understand the labeling, etc. the cost will be commensurate.</p>
3.b. Data Documentation	<p>Documentation</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,</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</li> </ul>	<p>Depending on the complexity, this may be completed by anyone from a Graduate Assistant under stipend to a salaried Principal Investigator</p>

<sup>8</sup> U.K Data Archive. (2022). *Costing data management for your research*. Retrieved from <https://ukdataservice.ac.uk/learning-hub/research-data-management/plan-to-share/costing/>

	processed, and quality controlled?	documentation is kept well-organized during research - low or no additional cost <ul style="list-style-type: none"> <li>• If documentation to be written or compiled specifically afterward - higher cost</li> </ul>	
4.a. Data Storage and Backup	Data Backup <ul style="list-style-type: none"> <li>• Does the institution provide regular backup or not?</li> <li>• Consider how frequently backups should be done, and how many backups should be stored.</li> <li>• Consider a secure backup, quality control, review schedule to ensure backups are being completed correctly</li> </ul>	<ul style="list-style-type: none"> <li>• Institutional backup - included in standard indirect costs/overheads</li> <li>• Additional backup needed - Cost according to number of copies to be kept, frequency of backup, and storage media need</li> <li>• Data recreation needed after failed or inadequate backup - potentially very high cost if possible at all. It is highly recommended you design data collection and storage with proper backups from the beginning</li> </ul>	<p>The Office of Information Technology provides free virtual servers to faculty and researchers. They can provide 25TB of data storage space.<sup>9</sup> This server space may not meet all funder security levels (See 5.b.)</p> <p>OIT has virtual servers based on Cisco UCS blade server platforms, VMware ESX platform software and NetApp SAN storage. The platform is built with redundancy and is auto correcting.<sup>10</sup></p>
4.b. Data Storage and Backup	Data Storage <ul style="list-style-type: none"> <li>• How much data storage space is needed for the entire duration of the</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> </ul>	<p>Research Computing can consult on data storage needs to determine whether institutional resources are</p>

<sup>9</sup> Research Computing. (2022). *Research Technology Support Services*. Retrieved from <https://www.boisestate.edu/rcs/research-technology-services/>. See also *Data Management: Storage, Sharing, and Publishing*. (2022). <https://www.boisestate.edu/rcs/data-management-storage-sharing-and-publishing/> and note researchers are encouraged to contact Research Computing ([researchcomputing@boisestate.edu](mailto:researchcomputing@boisestate.edu)) should they need more storage.

<sup>10</sup> Research Computing. (2022). *Grant Support*. Retrieved from <https://www.boisestate.edu/rcs/grant-support/>

	<p>project?</p> <ul style="list-style-type: none"> <li>Do you need to set up a data model and accompanying data?</li> </ul>	<ul style="list-style-type: none"> <li>If additional storage needed - cost server/disk space, as well as the cost of setting up and maintenance will need to be discussed with OSP</li> <li>Do you need a data warehouse or a database architect?</li> </ul>	<p>available and provide adequate space</p> <p>An IT Database Analyst or Data Coordinator could consult on data storage needs</p>
5.a. Data Access and Security	<p>Data Access</p> <ul style="list-style-type: none"> <li>Do external people require access to research data?</li> </ul>	<ul style="list-style-type: none"> <li>Does remote access via VPN or secure FTP need to be arranged for external people?</li> </ul>	<p>OIT and Research Computing can assist with existing services</p>
5.b. Data Access and Security	<p>Data Security</p> <ul style="list-style-type: none"> <li>Is there an institutional server available where you can store your data safely?</li> <li>Protect data from unauthorized use or from disclosure.</li> </ul>	<ul style="list-style-type: none"> <li>For confidential or privacy sensitive data, determining conditions for controlling access to data will require extra time and discussion with the OSP and ORC <ul style="list-style-type: none"> <li>Never assume that resources are available as indirect costs without consulting with OIT</li> </ul> </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>	<p>OIT and Research Computing can assist with free and existing services dependent upon the type of data being collected and the risk associated with that data.<sup>11</sup> Discuss data security <i>before</i> the collection phase to ensure all data are properly protected.</p>
6. Data Preservation and Archiving	<p>File Format</p> <ul style="list-style-type: none"> <li>Do data need to be converted to a standard or open format with long-term</li> </ul>	<ul style="list-style-type: none"> <li>For the purposes of long-term preservation and higher likelihood of usable formats in the future, using open file formats is highly</li> </ul>	<p>Likely completed by a Graduate Assistant under stipend or Student Research Technician 3</p>

<sup>11</sup> OIT Governance, Risk & Compliance. (2022). *Boise State University Data Classification Standards*. Retrieved from <https://www.boisestate.edu/oit-itgrc/it-standards-category/boise-state-university-data-classification-standard-2/>

	<p>validity for long-term preservation?</p>	<p>recommended<sup>12</sup></p> <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>	<p>Complex conversions may require a PI or co-PI at a commensurate cost</p>
<p>7.a. Data Sharing and Reuse</p>	<p>Anonymization</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>• Anonymization needs to be consistent throughout data collection and reflect the IRB approval (if applicable)</li> <li>• Both direct and indirect identifiers should be considered for removal or redaction<sup>13</sup></li> </ul>	<ul style="list-style-type: none"> <li>• If anonymization is planned before data collection or transcription/digitization - cost can be lowered</li> <li>• For audio-visual data - anonymization/editing voices or faces can be very costly and could reduce the usefulness of the data</li> <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.</li> </ul>	<p>Quantitative and Qualitative data can usually be anonymized by a Graduate Assistant under stipend, Student Research Technician 1, 2, or 3 at a salary rate of S2, S3, or S4 or by the research team at a commensurate cost</p> <p>Audio-visual anonymization may require a Videographer/Editor</p> <p>Many Boise State Offices (IRB, Tech Transfer, Albertsons Library, OSP) can</p>

<sup>12</sup> Library of Congress. (2020-2021). *LOC Recommended Formats Statement*. Retrieved from <https://www.loc.gov/preservation/resources/rfs/RFS%202020-2021.pdf>

<sup>13</sup> ICPSR. (2022). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (6th Edition). Retrieved from <https://www.icp.sr.umich.edu/files/deposit/dataprep.pdf> (See page 42)

		<p>interview transcripts) - costs can be reduced if anonymization is carried out during transcription (or at least highlighted/coded during transcription)</p> <ul style="list-style-type: none"> <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; pseudonymization will require more time</li> <li>• For files received of participants, check file properties and edit to remove disclosive information such as editor/author name</li> </ul>	<p>advise on data anonymization, but responsibility ultimately lies with the PI to ensure the data is properly deidentified.</p>
7.b. Data Sharing and Reuse	<p>Copyright</p> <ul style="list-style-type: none"> <li>• Do you need to seek copyright clearance before sharing data?<sup>14</sup></li> <li>• Do other parties hold copyright in the 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>Researchers should consult with Technology Transfer, OSP or General Counsel to ensure copyright approval. These consultations may take several business days to several weeks depending on the complexity of the issue.</p>
7.c. Data Sharing and Reuse	<p>Data Sharing</p> <ul style="list-style-type: none"> <li>• Will your data be deposited with a data center or institutional repository?</li> </ul>	<ul style="list-style-type: none"> <li>• How long is the data required to be available?</li> <li>• A public repository/data center/journal can provide you with</li> </ul>	<p>Completing a data repository upload form may take between 15 minutes to 4 hours or more, depending on</p>

<sup>14</sup> It is recommended that you review Boise State University *Policy #1090: Intellectual Property*. (2015). Retrieved from <https://www.boisestate.edu/policy/governance-legal/intellectual-property/> Researchers often do not realize the relationship between their datasets and the University.

	<ul style="list-style-type: none"> <li>• Which requirements exist to prepare data to particular standards, e.g. regarding documentation or format?</li> <li>• Do structured metadata need to be created when data are shared via a data center or archive, e.g. completing a deposit form for the ICPSR repository?</li> <li>• What data will be retained and what will not?</li> </ul>	<p>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</p> <ul style="list-style-type: none"> <li>• Data centers will have their own metadata forms. Consider using these beforehand</li> </ul>	<p>the complexity of the data and the upload form.</p> <p>ScholarWorks is a free data repository open to Boise State researchers, but can only handle non-sensitive data. Consult re3data.org<sup>15</sup> for disciplinary and generalist repositories that fit specific data criteria</p> <p>Costs vary depending on many factors, and range from free to several hundreds of dollars.</p> <p>Example: Dryad Digital Repository<sup>16</sup>: \$120 for the first 50GB, \$50 for each additional 10GB</p>
7.d. Data Sharing and Reuse	<p>Data Cleaning</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</li> </ul>	<ul style="list-style-type: none"> <li>• Data cleaning takes time</li> <li>• If carried out as part of data entry and preparation before data analysis - low additional cost</li> <li>• If needed afterwards - incurs</li> </ul>	<p>Some large data service companies claim to provide effective data cleaning/scrubbing for as low as \$6/hour or \$960 FTE<sup>17</sup>,</p>

<sup>15</sup> There are several generalist and disciplinary specific repositories that accept data deposits. Consult the Registry of Research Data Repositories for specific repositories matching specific deposit requirements. <https://www.re3data.org/>

<sup>16</sup> Dryad Digital Repository is accepted by the NIH as a generalist repository. NIH National Library of Medicine. (2022). *Generalist Repositories*. Retrieved from [https://www.nlm.nih.gov/NIHbmic/generalist\\_repositories.html](https://www.nlm.nih.gov/NIHbmic/generalist_repositories.html)

<sup>17</sup> Flatworld Solutions. (2022). *Data entry services pricing structure*. Retrieved from <https://www.flatworldsolutions.com/data-management/data-management-pricing.php>

	<p>used, check for anomalous values?</p> <ul style="list-style-type: none"> <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> <li>• Do you need to combine your data with other datasets for your research?</li> </ul>	higher costs	<p>however other professionals indicate that a competent data cleaning may cost \$90-\$150/hour<sup>18</sup></p> <p>A Data Coordinator may be able to complete this task</p>
7.e. Data Sharing and Reuse	<p>Digitization</p> <ul style="list-style-type: none"> <li>• Do analogue or paper-based research data (maps, newspaper clippings, photographs, images, text) need to be digitized to increase their potential 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 (OCR) 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 digitize tabular data - may be high cost</li> </ul>	<p>Albertsons Library does not provide digitization of research data</p> <p>Most commercial digitization services charge by the page e.g. \$0.10/page or a bulk rate \$350/1,000 pages</p>
8.a. General	<p>Roles and Responsibilities</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>Travel costs may be eligible as a direct cost if they are “necessary” and “reasonable and consistent with University’s documented</p>

<sup>18</sup> Du Soleil, Oz. (2013). *What does it cost to clean my data? (capturing Cerberus)*. Retrieved from <http://datascope.net/cost-of-data-cleansing/>

			travel policies.” Entertainment costs are not permitted. <sup>19</sup>
8.b. General	<p>Operationalizing Data Management</p> <ul style="list-style-type: none"> <li>• What measures are needed to implement and operationalize 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 through 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> <li>• Do you need a dedicated project manager?</li> </ul>	<p>Positions that may accomplish many data management tasks:</p> <ul style="list-style-type: none"> <li>• Data Coordinator</li> <li>• Research Analyst</li> <li>• Project Coordinator or Project Manager</li> </ul>

<sup>19</sup> Boise State University. (2017). *Basics of charging costs to a sponsored project*. Retrieved from <https://www.boisestate.edu/research-osp/charging-costs/>



This document is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License and derived from Grevatt, H. (2018). *Boise State University Research Data Management Costing Guide*.

Please cite as:

Grevatt, H. & Shook, E. (2022). *Boise State University Research Data Management Costing Guide*.

Attribution to the original content creators:

A. Westerhof, T.E. Pronk, A. van der Kuil, A. Mordant (2016). *Data Management Cost Guide*.

[https://www1.edugroepen.nl/sites/RDM\\_platform/Financieel1/Data%20Management%20Costs.aspx](https://www1.edugroepen.nl/sites/RDM_platform/Financieel1/Data%20Management%20Costs.aspx)

UK Data Service. (2013). *Data Management Costing Tool*. UK Data Archive, University of Essex.