**Data Management Planning Workbook**

Remember that a data management plan is a living document and should be reviewed and updated regularly. Use the sections in this notebook to document the management of your data in your project or lab. Not all sections may be necessary for your specific research or project.

The recommended structure for this data management workbook is as follows:

1. Description of project or research
   1. Project/Research summary
   2. Facility or lab group
   3. Funding Sources
   4. Dates of DM Workbook Versions
   5. PI / Researcher Contact
2. Data sources (Existing data)
3. Data collection, creation and analysis
   1. Types of data and file formats
   2. Data analysis and transformations
   3. Data organization methods
   4. Data quality control
4. Data administration
   1. Responsibilities
   2. Related policies
      1. Funder requirements
      2. Institutional / Department requirements
      3. Publisher requirements
   3. Data owners and stakeholders
   4. Privacy/Confidentiality
   5. Access and security
   6. Storage/Backups
5. Data sharing
   1. File formats, standards and conventions
   2. Sharing
6. Archiving
   1. Disposal
7. Data documentation and metadata
   1. Research project documentation
   2. Data files documentation
8. Budget
9. **Description of project or research**

Briefly summarize the type of study to help others understand the purposes for which the data are being collected or created.

* 1. **Project/Research summary**
* Abstract
  1. **Facility or lab group**
* Department name
* Lab name, institution
  1. **Funding sources**
* solicitation number(s) (if applicable)
* link(s) to your grant proposals and other documents
  1. **Dates of DM Workbook versions**
* Date of first DM Workbook version (YYYYMMDD)
* Date the DMP was last changed (YYYYMMDD)
  1. **PI / Researcher Contact**
* Contact Name, E-mail, Phone, Researcher ID (e.g. [ORCID](http://orcid.org/))

**2. Data sources (Existing data)**

While not required, it is good practice to see if there are existing data that could replace or augment the data you are planning to create.

* Have you searched the web and data archives for similar datasets?
* Are there any datasets that could assist with your research?
* How do the existing datasets fail to meet your requirements and therefore require new data to be created?

If you use existing datasets, please describe them using the following table, use one table for each existing dataset.

| **[*Dataset #]*** | **[*Name of Collection*]** |
| --- | --- |
| Description: | Describe the information that will be used and the nature and scale (e.g., national, regional, landscape, etc.) of the data. Include a link to the source of the existing data. |
| Format: | Identify the formats in which the data are maintained and made available. |
| Quality Checks: | Specify the procedures used to evaluate the existing data, including verification, validation, and an assessment of usability. |
| Source: | Identify the source for the data. |
| Restrictions: | Identify any limitations on access or reuse (e.g., sensitive data, restricted data, software with license restrictions, etc.) and provide justification for restriction. Provide citation or documentation describing limitations if due to policies or legal reasons. |
| Fees: | Identify any fees associated with acquiring the data. |
| Citation: | Provide citation for data product. If the data product can be found online, provide a URL. |

Data Inputs - Existing Collections: Based on the USGS - National Climate Change and Wildlife Science Center (NCCWSC) [DMP template](https://nccwsc.usgs.gov/sites/default/files/files/DMPv2dot1.docx): https://nccwsc.usgs.gov/sites/default/files/files/DMPv2dot1.docx

**3. Data collection, creation and analysis**

This section of the workbook deals with original data collected or generated during the research project. Also include here the data file organizations, transformations and analysis.

**3.1 Types of data and file formats**

|  |  |
| --- | --- |
| **Types of Data** | **File Formats during project** |
| **Observational** |  |
| * Captured in real-time * Usually irreplaceable * Examples: Sensor readings, telemetry, survey results, images |  |
| **Experimental** |  |
| * Data from lab equipment * Often reproducible, but can be expensive * Examples: gene sequences, chromatograms, magnetic field readings |  |
| **Simulation** |  |
| * Data generated from test models * Models and metadata, where the Input parameters are more important than output data * Reproducible with original code * Examples: climate models, economic models |  |
| **Derived or compiled** |  |
| * Reproducible (but very expensive) * Examples: text and data mining, compiled database, 3D models |  |

You should list all the data that will be created during the project. Use the following table to briefly describe your data files.

| **[*Data #]*** | **[*Provide a brief name to describe new data collection*]** |
| --- | --- |
| Description: | Describe the information that will be used and the nature and scale (e.g., national, regional, landscape, etc.) of the data that will be collected. |
| Data Management Resources: | Describe the proposal resources allocated for data management activities for the new data collected as a level of effort, total dollars allocated, or as a percentage of the total project’s cost. Resources could include people’s time or proposal funding. |
| Format: | Identify the formats in which the data will be generated and maintained. Consider “best formats” for sharing your data and describe in section 5 if this data is to be made available (shared). |
| Data Processing & Scientific Workflows: | Describe data processing steps or provide a scientific workflow you plan to use to manipulate the data, as appropriate. The new data should be described in a separate table. |
| Protocols: | Identify any standard protocols or methodologies that will be used to collect the data, if available. |
| Quality Checks: | Specify the procedures for ensuring data quality. (Can detail process in Section **3.4**) |
| Metadata: | Identify the metadata standard that will be used to describe the document (FGDC, ISO, EML, DDI etc.) |
| Volume Estimate: | Estimate the volume of information generated: megabyte (MB), GB, TB, or PB. |
| Access & Sharing: | During the project, specify who should have access to project information/products and what type of access (Public, Read, Write, No Access). |
| Citation: | Specify how the project’s data should be cited. |
| Contact: | Provide a point(s) of contact if questions arise related to the data and associated products (name, email, and phone number). |

Data Inputs - New Collections: Based on the USGS - National Climate Change and Wildlife Science Center (NCCWSC) [DMP template](https://nccwsc.usgs.gov/sites/default/files/files/DMPv2dot1.docx): https://nccwsc.usgs.gov/sites/default/files/files/DMPv2dot1.docx

**3.2 Data analysis and transformations**

| **[*Other Need #]*** | **[*Name of Software or Other Need*]** |
| --- | --- |
| Description: | Describe any software or other needs that are required for the project. Basic software such as Microsoft Office, Adobe, and an Internet Browser do not need to be provided. |
| Restrictions: | Identify any limitations on access or reuse that accompany the software or other needed items. |
| Fees: | Identify any fees or other costs associated with acquiring the software or other items. |
| Source/Link: | Provide a link or a source for the need if available. |

Software and other needs: Based on the USGS - National Climate Change and Wildlife Science Center (NCCWSC) [DMP template](https://nccwsc.usgs.gov/sites/default/files/files/DMPv2dot1.docx): https://nccwsc.usgs.gov/sites/default/files/files/DMPv2dot1.docx

**3.3 Data organization methods**

Consider how the data will be organized during the project, thinking about naming conventions, version control and folder structures. These will usually not be included on a Data Management (Sharing) Plan for funder requirements.

When organizing your files it is important to consider establishing a scheme for directory and file name conventions. Having a scheme that is documented within a lab or for personal use facilitates the retrieval of documents because there is an established organization to the files and because the file names have descriptive elements that aid in their understandability. In addition to the benefits to current users of these files, a standard scheme is also important for the long-term usability of the files that are transferred to the archives. (Bentley Historical Library, University of Michigan: *Digital Recordkeeping Best Practices for Directory and File Naming*: <http://bentley.umich.edu/dchome/resources/filenaming.php>.)

**Organizing Files**

* Directory Structure/File Naming Conventions

|  |
| --- |
| Rules used to create file names & examples from UC Boulder Libraries. *Guidelines on File Naming Conventions for Digital Collections*. <http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf>.  Rules   1. File name is 8‐9 characters long with a three digit file extension. 2. File name consists of numbers and lower case letters only. 3. The base file name is derived from the city, date, and page number printed on the  map. 4. The city is a three letter code created by the Map Library specifically for this digital  collection. 5. The date is the last two digits of the year in which the map was drawn. The date is  printed on the map. 6. The page number is printed in the upper right corner of the map sheet. The page  number in the file name is a three digit number. “001”, “002” ... “021”, “022” ... 7. Some maps include one or more unnumbered sheets. These files are denoted with  “un1”, “un2”, etc. in place of a three digit page number. 8. Revised map sheets have the same city, date, and page number as the original. To  maintain unique file names, an “r” for revised is placed between the date and page number on all revised map sheets.   Structure:  cccyy[r]ppp.ext, where:  ccc = City code designated by the Map Library  yy = Last two digits of the year printed on the map  r = The letter “r” designates revised maps. [Optional]  ppp = Three digit page number printed in the upper right corner of the sheet or “un” + sequence number for unnumbered sheets  Examples:  asp98003.tif  Map of Aspen, 1898, numbered sheet 3  cos07un3.tif  Map of Colorado City, 1907, third unnumbered sheet  sal86r001.tif  Revised map of Salida, 1886, numbered sheet 1A |

* File Naming Conventions for your Specific Discipline

(for examples, *Best Practices for Preparing Environmental Data Sets to Share and Archive, available:* <http://daac.ornl.gov/PI/BestPractices-2010.pdf>

* File Version Control
  + Software used (URL to software site)
  + OR File naming conventions for version control
    - FNL = final
    - DFT = draft
    - 2.0 or V2 = version of document

**3.4 Data quality control**

Describe quality assurance processes you will be using. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardized data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies or controlled codes. Here are some ways to control for data entry, collection errors:

* + Manually check 5 – 10% of data records (use a random representative sample)
  + Check for out-of-range values (plotting)
  + Display Location Data on a map
  + Use a data entry program
    - Program to catch typing errors
    - Program pull-down menu option

**4. Data administration**

**4.1 Responsibilities**

List who will be primarily responsible for implementing the DMP. Also note who is responsible for reviewing and modifying the data management plan.

In addition to the primary researcher(s), there might be others involved in the research process that take part in aspects of data management. Roles and responsibilities should be clearly defined, rather than assumed; this is especially important for collaborative projects that involve many researchers, institutions, and/or groups.

Using the activities in this workbook (each section heading), record who is responsible for each. As persons change, don’t remove or replace, but add the new person and close the date for the previous one. List the following for each person:

* Name
* E-mail
* Dates responsible
* Data Management Roles

|  |
| --- |
| **Examples of roles in data management**:   * data collector * metadata generator * data analyzer * project director * data model and/or database designer * computing staff responsible for backup and/or storage * staff responsible for running instruments * administrative support staff responsible for grant submission * specialized skills as defined in the plan (GIS, relational database design/implementation, computer programming of sensors/input forms, etc) * external data center or archive   *DataOne Best Practices*: *Define roles and assign responsibilities for data management*. <http://www.dataone.org/best-practices/define-roles-and-assign-responsibilities-data-management>. |

**4.2 Related policies**

List any relevant policies. Some policies (such as data archiving) are relevant to all research projects, whereas privacy (see section **4.4**) will usually be associated with medical and social science projects.

* Does any of your data contain personal information that must be kept confidential?
* Does your funding agreement require data archiving?
* Does your funder have a Research Data Management policy?

**4.2.1 Funding requirements**

* funder links (for data management, data sharing, data archiving, data policy guidelines)
* funder special requirements per solicitation or award instructions

**4.2.2 Institutional / Department requirements**

* Does your institution and/or department have data management guidelines?
  + Institutional links (for data management, data sharing, data policy guidelines)
  + Departmental links (for data management, data sharing, data policy guidelines)

**4.2.3 Publisher requirements**

* Publisher links (for data management, data sharing, data policy guidelines)
* Author guidelines for making available research data which supports results in the paper

**4.3 Data owners and stakeholders**

List the owners and stakeholders of the data. Include links to data retention policies. Are there any restrictions on the reuse of third-party data (usage rights of existing data).

|  |
| --- |
| **Example polices from UVa:**   * UVa Data Rights & Responsibilities   <http://dmconsult.library.virginia.edu/data-rights-and-responsibilities-guidance-1-0/>   * UVa Laboratory Notebook and Recordkeeping Policy <https://policy.itc.virginia.edu/policy/policydisplay?id=RES-002> |

**4.4 Privacy/Confidentiality**

Managing ethical concerns may include: anonymization of data, referring to your Institutional Review Board (IRB) protocol, or formal consent agreements.

* Have you gained consent for data sharing, some or all of the data?
* How will you protect the identity of participants?
* How will sensitive data be handled to ensure it is stored and transferred securely?

|  |
| --- |
| **Policy Examples from UVa:**   * Read the Institutional Review Board for Health Sciences Research requirements <http://www.virginia.edu/vpr/irb/hsr/index.html> * Read the Institutional Review Board for Social & Behavioral Sciences requirements <http://www.virginia.edu/vpr/irb/sbs/> |

**4.5 Access and security**

List who will have access to the research data and what access permissions they will have for specific data. If the data will be distributed at some point, list the access restrictions and any embargoes that will be used.

Describe how the access permissions will be enforced and what IT security practices will be used. Outline any appropriate security measures and note any formal standards that you will comply with. Do you or your institution have a data protection or security policy that you must follow? Security measures might be defined in your IRB.

**Security**

Network security

* Keep confidential data off the Internet
* Put sensitive materials on computers not connected to the internet

Physical Security

* Restrict access to buildings and rooms where computers or media are kept
* Only let trusted individuals troubleshoot computer problems

Computer Systems & Files

* Virus protection software used / updated regularly
* Don't send confidential data via e-mail or FTP
* Use encryption, if you must
  + Include here instructions on encrypting and how to de-encrypt
* Use passwords on files and computers **(do not keep passwords in this workbook)**

**4.6 Storage/Backups**

Keep security in mind when doing backups.

**Storage**

* File formats
* Use [files formats](http://www2.lib.virginia.edu/brown/data/files.html) that will be useable in the long-term, not dependent on software version (see section **5.1** on **File formats, standards and conventions**)
* Media Types (CD, DVD, Tapes, etc.)
* Media Migration (Schedule)
* Environmental Conditions (specify where media, computers will be physically located)

**Backups**

List what data will be backed up and what the backup schedule is.

* Is there a backup service already available or will you need to do it yourself?
* How often will backups occur?
* Who will be responsible for performing backups?
* How will sensitive data be backed up?

Do you have sufficient storage or will you need to include costs for backup hardware and software?

Backup Specifics

* Software
* Responsibility (who)
* Schedule
* Locations (3 Places)
* Here \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* Near \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* Far \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* Media for backups
* Test your backups (date recovery tested)

**5. Data sharing**

**5.1 File formats, standards and conventions**

List what formats, standards, and conventions will apply to each data item. Justify the use of particular formats in terms of usability, longevity, and suitability for archiving.

* Will other researchers be able to use this format?
* Will this format be usable in ten years time?
* Does your archive accept this file format or can you easily convert to an acceptable format?
* Include formats used because of sustainability, and describe transformations needed

|  |  |
| --- | --- |
| **Things to consider:**   * Accessible in the future * Non-proprietary * Open, documented standard * Common, used by the research community * Standard representation (ASCII, Unicode) * Not specific toparticular software is a plus | **Best Formats:**   * Unencrypted * Uncompressed * PDF, not Word * ASCII, not Excel * MPEG-4, not QuickTime * TIFF or JPEG2000, not GIF or JPG * XML or RDF, not RDBMS |

|  |  |  |
| --- | --- | --- |
| **Type of data** | **Acceptable formats for sharing, reuse and preservation** | **Other acceptable formats for data preservation (for the UK Data Archive)** |
| **Quantitative tabular data with extensive metadata**  a dataset with variable labels, code labels, and defined missing values, in addition to the matrix of data | SPSS portable format (.por)  delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) containing metadata information  some structured text or mark-up file containing metadata information, e.g. DDI XML file | proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta)  MS Access (.mdb/.accdb) |
| **Quantitative tabular data with minimal metadata**  a matrix of data with or without column headings or variable names, but no other metadata or labeling | comma-separated values (CSV) file (.csv)  tab-delimited file (.tab)  including delimited text of given character set with SQL data definition statements where appropriate | delimited text of given character set - only characters not present in the data should be used as delimiters (.txt)  widely-used formats, e.g. MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf) and OpenDocument Spreadsheet (.ods) |
| **Geospatial data**  vector and raster data | ESRI Shapefile (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn)  geo-referenced TIFF (.tif, .tfw)  CAD data (.dwg)  tabular GIS attribute data | ESRI Geodatabase format (.mdb)  MapInfo Interchange Format (.mif) for vector data  Keyhole Mark-up Language (KML) (.kml)  Adobe Illustrator (.ai), CAD data (.dxf or .svg)  binary formats of GIS and CAD packages |
| **Qualitative data**  textual | eXtensible Mark-up Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml)  Rich Text Format (.rtf)  plain text data, ASCII (.txt) | Hypertext Mark-up Language (HTML) (.html)  widely-used proprietary formats, e.g. MS Word (.doc/.docx)  some proprietary/software-specific formats, e.g. NUD\*IST, NVivo and ATLAS.ti |
| **Digital image data** | TIFF version 6 uncompressed (.tif) | JPEG (.jpeg, .jpg) but only if created in this format  TIFF (other versions) (.tif, .tiff)  Adobe Portable Document Format (PDF/A, PDF) (.pdf)  standard applicable RAW image format (.raw)  Photoshop files (.psd) |
| **Digital audio data** | Free Lossless Audio Codec (FLAC) (.flac) | MPEG-1 Audio Layer 3 (.mp3) but only if created in this format  Audio Interchange File Format (AIFF) (.aif)  Waveform Audio Format (WAV) (.wav) |
| **Digital video data** | MPEG-4 (.mp4)  motion JPEG 2000 (.mj2) |  |
| **Documentation and scripts** | Rich Text Format (.rtf)  PDF/A or PDF (.pdf)  HTML (.htm)  OpenDocument Text (.odt) | plain text (.txt)  some widely-used proprietary formats, e.g. MS Word (.doc/.docx) or MS Excel (.xls/.xlsx)  XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0 |

UK Data Archive. *Create & Mange Data: Formatting Your* Data. <http://www.data-archive.ac.uk/create-manage/format/formats-table>.

You may need to convert your data files to a preservation file format.

It is always a good idea to contact the manager(s) of the repository or archive where you intend to store your data to see if they have specific format requirements.

|  |  |
| --- | --- |
| **Types of Data** | **File Formats for sharing** |
| **Observational** |  |
| * Captured in real-time * Usually irreplaceable * Examples: Sensor readings, telemetry, survey results, images |  |
| **Experimental** |  |
| * Data from lab equipment * Often reproducible, but can be expensive * Examples: gene sequences, chromatograms, magnetic field readings |  |
| **Simulation** |  |
| * Data generated from test models * Models and metadata, where the Input parameters are more important than output data * Reproducible with original code * Examples: climate models, economic models |  |
| **Derived or compiled** |  |
| * Reproducible (but very expensive) * Examples: text and data mining, compiled database, 3D models |  |

**5.2 Sharing**

Decide which data to keep and for how long. This could be based on any obligations to retain certain data, the potential reuse value, and what is economically viable to keep. How will your data be shared (see Section **6** on Archiving for long-term preservation)?

* What data will be shared?
* What additional efforts are required to prepare the data for sharing?
* What facilities will be used/required to distribute the data?
* How will the data be licensed?
* What access restrictions will be placed on each item of data?

**6. Archiving**

Estimate the amount of storage space required for archiving and which archives or repositories you intend to use. Outline the plans for preparing and documenting data (metadata) for archiving. If the data is sensitive, describe how you will ensure the data will be safely disposed.

**Where to find archives to deposit your research data?**

* re3data (<http://www.re3data.org>) is a global registry of research data repositories.
* Databib (<http://databib.org>) is a tool to help identify and locate online repositories of research data (for archiving and for find existing data).
* NIH Data Sharing Repositories (<http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html>) is a list of NIH-supported data repositories that accepts submissions from NIH-funded investigators (and others).

Check repositories’ data submission policies to find what is required to submit data. Also check out to see if your institution has an institutional repository for research data.

* Which archive or repository will you use?
* How long must you keep your data archived?
* When do you plan to archive each item and will they have an embargo period?
* How much time and resources will be required for archiving?
* What metadata will be needed?

**6.1 Disposal**

* Does your unit, department or institution have a records retention policy or data destruction requirements?

|  |
| --- |
| **Example from UVa Records Management Policy** <https://policy.itc.virginia.edu/policy/policydisplay?id=IRM-017> |

**7. Data documentation and metadata**

Ensuring that data can be understood, interpreted and used requires clear and detailed data documentation. Sharing data for long-lasting usability would be impossible without documentation (also known as metadata).

Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. Wherever possible you should identify and use existing community standards.

**7.1 Research project documentation**

* Context of data collection
* Who created or contributed to the data and when
* Data collection methods
* Structure, organization of data files
* Data sources used (see Section **2**)
* Data validation, quality assurance
* Transformations of data from raw data through to the analysis stage
* Information on confidentiality, access & use conditions

**7.2 Data Files documentation**

* Variable names and descriptions
* Explanation of codes and classification schemes used (Codebooks)
* Algorithms used to transform data
* File format and software (including version) used

|  |
| --- |
| **ReadMe File Example**  This template is to be used as a template for the ReadMe file required for dataset (data package) deposit in Libra, UVa’s Institutional Repository (<http://libra.virginia.edu>).  A README file is required for each data package you deposit, and should be saved as a TEXT file named "README-{*Metadata/Dataset title*}-{*Todays date*}.txt" (ex. "ReadMe-BeetleData-20080131"). A ReadMe file must be submitted along with the dataset file(s). Multiple ReadMe files may be submitted if it is necessary to document each data file separately.  The outline below should be completed with information relevant to the submitted data package or data file:  Required information:   * 1. File names, directory structure (for zipped files) and brief description of each file or file type   2. Definitions of acronyms, site abbreviations, or other project-specific designations used in the data file names or documentation files, if applicable   3. Description of the parameters/variables (column headings in the data files) and units of measure for each parameter/variable, including special codes, variable classes, GIS coverage attributes, etc. used in the data files themselves, including codes for missing data values, if applicable: * column headings for any tabular data * the units of measurement used * what symbols/codes are used to record missing data * any specialized formats or abbreviations used * any additional related data collected that was not included in the current data package   1. Uncertainty, precision, and accuracy of measurements, if known   2. Environmental conditions, if appropriate (e.g., cloud cover, atmospheric influences, etc.)   3. Method(s) for processing data, if data other than raw data are being contributed   4. Standards or calibrations that were used   5. Specialized software (including version number) used to prepare, to analyze and/or needed to read the dataset, if applicable   6. Quality assurance and quality control that have been applied, if applicable   7. Known problems that limit the data's use or other caveats (e.g., uncertainty, sampling problems, blanks, QC samples)   8. Date dataset was last modified   9. Relationships with any ancillary datasets outside of this dataset, if applicable   Optional information:   * 1. Methodology for sample treatment and/or analysis, if applicable   2. Example records for each data file (or file type)   3. Files names of other documentation that are being submitted along with the data and that would be helpful to a secondary data user, such as pertinent field notes or other companion files, publications, etc. |

**8. Budget**

Now that the data management activities and responsibilities have been established, you can estimate the costs of data management for your project. Often the time involved in documenting, writing metadata, and archiving are underestimated.

The following chart includes cost activities which are part of a more comprehensive costing tool and check list developed by the UK Data Service. The UK Data Service – Data management costing tool and checklist with steps on how to use it is found here:

<http://www.data-archive.ac.uk/media/247429/costingtool.pdf>

|  |  |  |  |
| --- | --- | --- | --- |
| **Activity** | **Comments and Suggestions** | **✔** | **Cost** |
| **Data cleaning**   * Do quantitative data need to be cleaned, checked or verified before sharing, e.g. check validity of codes used, check for anomalous values? * Will data match documentation, e.g. same number of variables, cases, records, files? * Does textual information in data need to be spell-checked? | * If carried out as part of data entry and preparation before data analysis – low or no additional cost * If needed afterwards – higher cost |  |  |
| **Transcription**   * Will you transcribe qualitative data (e.g. recorded interviews of 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? * Is full or partial transcription needed? * Is translation needed? * Will you need to develop a standard transcription template or transcription guidelines, to ensure consistent formatting? | * If part of research practice – very low or not additional cost * If not planned as part of research practice – potentially high additional cost * Is additional hardware/software needed? * Consider cost of (time needed for) developing procedures, templates and guidance for transcribers * Calculate time needed for transcription – four to eight hours per recording (for a 1 hour recording) |  |  |
| **File Format**   * Do data need to be converted to a standard or open format with long-term validity for long-term preservation? | * Is additional hardware or software needed for conversion? * For audio-visual data, converting to open digital formats can be time-consuming or require special equipment and or software * For databases, conversion may require checking for truncation, loss of metadata or annotation, loss of relationships, etc. |  |  |
| **Data Storage**   * How much data storage space is needed for the entire duration of the project? * 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 | * If storage is provided by the institution – cost is included in standard indirect costs or overheads * If additional storage needed – cost of server/disk space, as well as the cost of setting up and maintenance costs |  |  |
| **Data Backup**   * Does the institution provide regular backup or not? * Consider how frequently backups should be done, how many backups should be stored | * Institutional backup – included in standard indirect costs (overhead costs) * Additional backup needed – cost according to number of copies to be kept, frequency of backup and storage media needed |  |  |
| **Operational Data Management**   * What measures are needed to implement and operationalize data management throughout the research lifecycle? | * Do you need extra time and resources to implement data management throughout your research, e.g. regular team meetings, setting up a collaborate research environment? * If staff training is required – higher cost * Do you need a dedicated data manger? |  |  |

UK Data Archive. *UK Data Service – Data management costing tool and checklist.* <http://www.data-archive.ac.uk/media/247429/costingtool.pdf>.

**For Further Reading**

Australian National Data Service. (2011). *Metadata Guide Working Level.* Available online: <http://www.ands.org.au/guides/metadata-working.html>.

Australian National University. (2013). *ANU Data Management Manual: Managing Digital Research Data at the Australian National University* (6TH ed.). Canberra, AU. Available online: <http://anulib.anu.edu.au/_resources/training-and-resources/guides/DataManagement.pdf>.

Digital Curation Center. (2013). *Checklist for a Data Management Plan*. v.4.0. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/data-management-plans>.

Digital Curation Center. (2013). *Disciplinary Metadata*. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/metadata-standards>.

Hook, Les A., Suresh K. Santhana Vannan, Tammy W. Beaty, Robert B. Cook, and

Bruce E. Wilson. 2010. *Best Practices for Preparing Environmental Data Sets to Share and Archive*. Available online (<http://daac.ornl.gov/PI/BestPractices-2010.pdf>) from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/BestPractices-2010.

Inter-university Consortium for Political and Social Research (ICPSR). (2012). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (5th ed.). Ann Arbor, MI. Available online: <http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>.

MIT Libraries. (2013). *Data Management and Publishing*. <http://libraries.mit.edu/guides/subjects/data-management/index.html>.

UK Data Archive. (2013). *Create & Manage Data.* University of Essex. Available online: <http://www.data-archive.ac.uk/create-manage>.

USGS. (2013). *USGS Data Management*. Available online: <http://www.usgs.gov/datamanagement/>.

Van den Eynde, Veerle, et al. (2011). *Managing and Sharing Data: Best Practice for Researchers* (3rd ed.). Colchester, UK: UK Data Archive. Available online: <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>.

Data Management Planning Workbook by [UVa Library Data Management Consulting Group](http://dmconsult.library.virginia.edu) is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit: <http://creativecommons.org/licenses/by-sa/4.0/>.