Data Management Plan Template:
Mixed Methods (Surveys & Qualitative Research)

Abstract

This mixed methods data management plan template was developed for use within the Portage DMP Assistant and is meant to assist researchers conducting mixed methods research (e.g., surveys and qualitative interviews/focus groups) in developing high quality data management plans to support their research. The template consists of a number of categories, questions, and customized guidance that relate directly to mixed methods research.

Administrative Details

Template Author(s): James Doiron, University of Alberta

Published: March 2, 2021

DOI: 10.5281/zenodo.4573319

Contact: Portage Network - portage@carl-abrc.ca, portagenetwork.ca

Version:

Version	Date	Changes
1.0	2021-03-02	Formatted for inaugural publication.

Research Data Management Policies

Are there any research data management policies in place that outline requirements and/or best practice guidance regarding the management of your data? If so, provide details and, if helpful, URL links to these policies.

Examples of research data management policies that may be in place include those set forth by funders, post secondary institutions, legislation, and communities.

Examples of these might include:

- Tri-Agency Statement of Principles on Digital Data Management
- First Nations Information Governance Centre Principles of OCAP
- Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS-2)

Data Collection

Describe the type(s) of data that you will collect, including all survey, interview and/or focus group data. If there are any additional types of data that will be collected or generated describe these as well.

Having a clear understanding of all the data that you will collect or use within your project will help with planning for their management.

Include a general description of each type of data related to your project, including the formats that they will be collected in, such as audio or video files for qualitative interviews and focus groups and survey collection software or file types.

As well, provide any additional details that may be helpful, such as the estimated length (number of survey variables/length of interviews) and quantity (number of participants to be interviewed) both of surveys and interviews.

Are there any existing data that you can re-use and that will provide insight or answer any of your research questions? If so, please explain how you will obtain these data and integrate them into your research project.

There are many potential sources of existing data, including research data repositories, research registries, and government agencies.

Example of these include:

- Statistics Canada Public Use Microdata Files available through the Data Liberation Initiative (DLI)
- The Canadian based Federated Research Data Repository (FRDR)
- Research data repositories, such as those listed at re3data.org
- Data repositories that house qualitative research data such as the Qualitative Data Repository (QDR), Dataverse, and the UK Data Archives QualiBank

You may also wish to contact the Library at your institution for assistance in searching for any existing data that may be useful to your research.

It is important to identify and understand as early as possible the methods which you will employ in collecting your data to ensure that they will support your needs, including supporting the secure collection of sensitive data if applicable.

Describe the method(s) that you will use to collect your data.

Include a description of any methods that you will use to collect data, including electronic platforms or paper based methods. For electronic methods be sure to include descriptions of any privacy policies as well as where and how data will be stored while within the platform.

For an example of a detailed mixed methods description, see this Portage DMP Exemplar in either English or French.

There are many electronic survey data collection platforms to choose from (e.g., Qualtrics, REDCap, Hosted in Canada Surveys). Understanding how and where your survey data will be collected and stored is an essential component of managing your data and ensuring that you are adhering to any security requirements imposed by funders or research ethics boards.

Additionally, it is important to clearly understand any security and privacy policies that are in place for any given electronic platform that you will use for collecting your data - examples of such privacy policies include those provided by Qualtrics (survey) and Zoom (interviews).

If interview and/or focus group audio recordings will be transcribed, describe how this will securely occur, including if it will be performed internally to the research team or externally (outsourced), and/or if any software and/or electronic platforms or services will be used for transcribing.

To support transcribing activities within your research project, it is recommended that you implement a transcribing protocol which clearly outlines such things as formatting instructions, a summary of contextual metadata to include, participant and interviewer anonymization, and file naming conventions.

When outsourcing transcribing services, and especially when collecting sensitive data, it is important to have a confidentiality agreement in place with transcribers, including a protocol for their deleting any copies of data once it has been transcribed, transferred, and approved. Additionally, you will need to ensure that methods for transferring and storing data align with any applicable funder or institutional requirements.

Describe how your data will be securely transferred, including from data collection devices/platforms and, if applicable, to/from transcriptionists.

Transferring of data is a critical stage of the data collection process, and especially so when managing sensitive information. Data transfers may occur:

- from the field (real world settings)
- from data providers
- between researchers
- between researchers & stakeholders

It is best practice to identify data transfer methods that you will use before your research begins.

Some risks associated with the transferring of data include loss of data, unintended copies of data files, and data being provided to unintended recipients. You should avoid transferring data using unsecured methods, such as email. Typical approved methods for transferring data include secure File Transfer Protocol (SFTP), secure extranets, or other methods approved by your institution.

Talk to your local IT support to identify secure data transferring methods available to you.

Describe all of the file formats that your data will exist in, including for the various versions of both survey and qualitative interview/focus group data. Will these formats allow for data re-use, sharing and long-term access to the data?

Ensuring that your data files exist in non-proprietary formats helps to ensure that they are able to be easily accessed and reused by others in the future.

Examples of non-proprietary file formats include:

Surveys: CSV; HTML; Unicode Transformation Formats

Qualitative interviews:

- Audio files→ MP3; FLAC; Ogg
- Video files→ MP4; .mkv
- Transcriptions→ Plain Text, such as ASCII; CSV; HTML

For more information and resources pertaining to file formats you may wish to visit:

- Library & Archives Canada
- Data Curation Network
- UK Data Service

Documentation and Metadata

Describe any documentation and metadata that will be used in order to ensure that data are able to be read and understood both during the active phases of the project and in the future.

Include a description of the survey codebook(s) (data dictionary), as well as how it will be developed and generated. You should also include a description of the interview data that will be collected, including any important contextual information and metadata associated with file formats.

Your documentation may include study-level information about:

- who created/collected the data
- when it was created
- any relevant study documents
- conditions of use
- contextual details about data collection methods and procedural documentation about how data files are stored, structured, and modified.

A complete description of the data files may include:

- naming and labelling conventions
- explanations of codes and variables
- any information or files required to reproduce derived data.

More information about both general and discipline specific data documentation is available at https://www.dcc.ac.uk/guidance/standards/metadata.

Describe the file naming conventions that will be used in order to support quality assurance and version-control of your files and to help others understand how your data are organized.

For guidance on file naming conventions please see the University of Edinburgh.

Describe how you will ensure that documentation and metadata are created, captured and, if necessary, updated consistently throughout the research project.

High quality documentation and metadata help to ensure accuracy, consistency, and completeness of your data. It is considered best practice to develop and implement protocols that clearly communicate processes for capturing important information throughout your research project. Example topics that these protocols might cover include file naming conventions, file versioning, folder structure, and both descriptive and structural metadata.

Researchers and research staff should ideally have the opportunity to contribute to the content of metadata protocols, and it is additionally useful to consult regularly with members of the research team to capture any potential changes in data collection/processing that need to be reflected in the documentation.

Describe any metadata standard(s) and/or tools that you will use to support the describing and documenting of your data.

Metadata are descriptions of the contents and context of data files. Using a metadata standard (a set of required fields to fill out) helps to ensure that your documentation is consistent, structured, and machine-readable, which is essential for depositing data in repositories and making it easily discoverable by search engines.

There are both general and discipline-specific metadata standards and tools for research data.

One of the most widely used metadata standards for surveys is DDI (Data Documentation Initiative), a free standard that can document and manage different stages in the research data lifecycle including data collection, processing, distribution, discovery and archiving.

For assistance with choosing a metadata standard, support may be available at your institution's Library or contact support@portagenetwork.ca.

Storage, Backup, and Access

Describe where, how, and for how long data will be securely stored during the *active phases* of the research project. If any data are to be collected through the use of electronic platforms, account for their usage within your data storage description. Include a description of any policies and procedures that will be in place to ensure that data are regularly backed-up.

Data storage is a critical component of managing your research data, and secure methods should always be used, especially when managing sensitive data. Storing data on USB sticks, laptops, computers, and/or external hard drives without a regular backup procedure in place is not considered to be best practice due to their being a risk both for data breaches (e.g., loss, theft) as well as corruption and hardware failure. Likewise, having only one copy, or multiple copies of data stored in the same physical location does little to mitigate risk.

Many universities offer networked file storage which is automatically backed up. Contact your local (e.g., faculty or organization) and/or central IT services to find out what secure data storage services and resources they are able to offer to support your research project.

Additionally, you may wish to consider investigating Compute Canada's Rapid Access Service which provides Principal Investigators at Canadian post-secondary institutions with a modest amount of storage and cloud resources at no cost.

Describe how members of the research team will securely access and work with data during the active phases of the research project.

It is important to determine at the early stages of your research project how members of the research team will appropriately access and work with data. If researchers will be working with data using their local computers (work or personal) then it is important to ensure that data are securely transferred (see previous question on data transferring), computers may need to be encrypted, and that all processes meet any requirements imposed by funders, institutions, and research ethics offices.

When possible, it can be very advantageous to use a cloud-based environment so that researchers can remotely access and work with data, reducing the need for data transferring and associated risks, as well as unnecessary copies of data existing.

One such cloud environment that is freely available to Canadian researchers is Compute Canada's Rapid Access Service.

Describe how much storage space you will require during the active phases of the research project, being sure to take into account file versioning and data growth.

Think about all of the data that will be generated, including their various versions, and estimate how much space (e.g., megabytes, gigabytes, terabytes) will be required to store them.

The type of data you collect, along with the length of time that you require active storage, will impact the resources that you require. Textual and tabular data files are usually very small (a few megabytes) unless you have a lot of data. Video files are usually very large (hundreds of megabytes up to several gigabytes). If you have a large amount of data (gigabytes or terabytes), it will be more challenging to share and transfer it. You may need to consider networked storage options or more sophisticated backup methods.

You may wish to contact your local IT services to discuss what data storage options are available to you, or consider the use of Compute Canada's Rapid Access Service.

Preservation

Describe how you will ensure that your data is preservation ready, including the file format(s) that they will be preserved in and. Explain how you will prevent data from being lost while processing and converting files.

Proprietary data formats are not optimal for long-term preservation of data as they typically require specialized licensed software to open them. Such software may have costs associated with its use, or may not even be available to others wanting to re-use your data in the future.

Non-proprietary file formats, such as comma-separated values (.csv), text (.txt) and free lossless audio codec (.flac), are considered preservation-friendly. The UK Data Archive provides a useful table of file formats for various types of data. Keep in mind that preservation-friendly files converted from one format to another may lose information (e.g. converting from an uncompressed TIFF file to a compressed JPG file), so changes to file formats should be documented.

Identify the steps required to ensure the data you are choosing to preserve is error-free, and converted to recommended formats with a minimal risk of data loss following project completion. Some strategies to remove identifiers in images, audio, and video (e.g. blurring faces, changing voices) also remove information of value to other researchers.

See this Portage DMP Exemplar in English or French for more help describing preservation-readiness.

Describe where you will preserve your data for long-term preservation, including any research data repositories that you may be considering to use. If there are any costs associated with the preservation of your data, include those details.

A research data repository is a technology-based platform that allows for research data to be:

- Deposited & described
- Stored & archived
- Shared & published
- Discovered & reused

There are different types of repositories including:

- Proprietary (paid for services)
- Open source (free to use)
- Discipline specific

A key feature of a trusted research data repository is the assignment of a digital object identifier (DOI) to your data - a *unique persistent identifier* assigned by a registration agency to identify digital content and provide a persistent link to its location, enabling for long-term discovery.

Dataverse is one of the most popular research data repository platforms in Canada for supporting the deposition of survey data and qualitative text files. Key features of Dataverse include the assignment of a DOI, the ability to make your data both open or restricted access, built in data citations, file versioning, and the ability to create customized terms of use pertaining to your data. Contact your local university Library to find out if there is a Dataverse instance available for you to use.

Re3data.org is an online registry of data repositories, which can be searched according to subject, content type and country. Find a list of Canadian research data repositories.

Sharing and Reuse

Describe what data you will be sharing, including which version(s) (e.g., raw, processed, analyzed) and in what format(s).

Consider which data you are planning to share or that you may need to share in order to meet funding or institutional requirements. As well, think about which data may possibly be restricted for reasons relating to confidentiality and/or privacy. If you are planning to share either/both survey and qualitative interviews data that require de-identification, explain how any necessary direct and indirect identifiers will be removed.

Examples of file versions are:

- **Raw**: Original data that has been collected and not yet processed or analysed. For surveys this will be the original survey data, and for qualitative interviews this will most often be the original audio data as well as raw transcriptions which are verbatim copies of the audio files.
- **Processed:** Data that have undergone some type of processing, typically for data integrity and quality assurance purposes. For survey data, this may involve such things as deletion of cases and derivation of variables. For qualitative interview data, this may involve such things as formatting, and de-identification and anonymization activities.
- **Analyzed:** Data that are already processed and have been used for analytic purposes. Both for surveys and qualitative interviews, analyzed data can exist in different forms including in analytic software formats (e.g. SPSS, R, Nvivo), as well as in text, tables, graphs, etc.

Remember, research involving human participants typically requires participant consent to allow for the sharing of data. Along with your data, you should ideally include samples of the study information letter and participant consent form, as well as information relating to your approved institutional ethics application.

Describe whether there will be any restrictions placed on your data when they are made available and who may access them. If data are not openly available, describe the process for gaining access.

It may be necessary or desirable to restrict access to your data for a limited time or to a limited number of people, for:

- ethical reasons (privacy and confidentiality)
- economic reasons (patents and commercialization)
- intellectual property reasons (e.g. ownership of the original dataset on which yours is based)
- or to comply with a journal publishing policy.

Strategies to mitigate these issues may include:

- anonymising or aggregating data (see additional information at the UK Data Service or the Portage Network)
- gaining participant consent for data sharing
- gaining permissions to share adapted or modified data
- and agreeing to a limited embargo period.

If applicable, consider creating a Terms of Use document to accompany your data.

What type of end-user license will you include with your data?

Licenses determine what uses can be made of your data. Funding agencies and/or data repositories may have end-user license requirements in place; if not, they may still be able to guide you in the development of a license. Once created, it is considered as best practice to include a copy of your end-user license with your Data Management Plan. Note that only the intellectual property rights holder(s) can issue a license, so it is crucial to clarify who owns those rights.

There are several types of standard licenses available to researchers, such as the Creative Commons licenses and the Open Data Commons licenses. In fact, for most datasets it is easier to use a standard license rather than to devise a custom-made one. Note that even if you choose to make your data part of the public domain, it is preferable to make this explicit by using a license such as Creative Commons' CC0.

Read more about data licensing: UK Digital Curation Centre.

Responsibilities and Resources

Who will be responsible for data management during the project (i.e., during collection, processing, analysis, documentation)? Identify staff and organizational roles and their responsibilities for carrying out the data management plan (DMP), including time allocations and training requirements.

Research data management is a shared responsibility that can involve many research team members including the Principal Investigator, co-investigators, collaborators, trainees, and research staff. Some projects warrant having a dedicated research data manager position. Think about your project and its needs, including the time and expertise that may be required to manage the data and if any training will be required to prepare members of the research team for these duties.

Larger and more complex research projects may additionally wish to have a research data management committee in place which can be responsible for data governance, including the development of policies and procedures relating to research data management. This is a useful way to tap into the collective expertise of the research team, and to establish robust policies and protocols that will serve to guide data management throughout your project.

How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

It is important to think ahead and be prepared for potential PI and/or research team members changes should they occur. Developing data governance policies that clearly indicate a succession strategy for the project's data will help greatly in ensuring that the data continue to be effectively and appropriately managed. Such policies should clearly describe the process to be followed in the event that the Principal Investigator leaves the project. In some instances, a co-investigator or the department or division overseeing this research will assume responsibility.

What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

Estimate as early as possible the resources and costs associated with the management of your project's data. This estimate should incorporate costs incurred both during the active phases of the project as well as those potentially required for support of the data once the project is finished, including preparing the data for deposit and long-term preservation.

Many funding agencies will provide support for research data management, so these estimates may be included within your proposed project budget. Items that may be pertinent to mixed methods research include such things as a dedicated research data management position (even if it is part-time), support for the use of a digital survey data collection platform, computers/laptops, digital voice recorders, specialized software, transcription of qualitative interviews, data storage, data deposition, and data preservation.

Ethical and Legal Compliance

If applicable, what strategies will you undertake to address secondary uses of data, and especially those which are sensitive in nature?

Obtaining the appropriate consent from research participants is an important step in assuring Research Ethics Boards that the data may be shared with researchers outside your project. The consent statement may identify certain conditions clarifying the uses of the data by other researchers, as well as what version(s) of the data may be shared and re-used. For example, it may stipulate that the data will only be shared for non-profit research purposes, that the data will not be linked with personally identified data from other sources, and that only de-identified and/or aggregated data may be reused. In the case of qualitative interviews, this may include only the de-identified transcriptions of interviews and/or analytic files containing de-identified contextual information.

Sensitive data in particular should always receive special attention and be clearly identified and documented within your DMP as to how they will be managed throughout your project including data collection, transferring, storage, access, and both potential sharing, and reuse.

Your data management plan and deposited data should both include an identifier or link to your approved research ethics application form as well as an example of any participant consent forms.

How will you manage legal, ethical, and intellectual property issues?

Compliance with privacy legislation and laws that may impose content restrictions in the data should be discussed with your institution's privacy officer, research services office, and/or research ethics office.

Include here a description concerning ownership, licensing, and intellectual property rights of the data. Terms of reuse must be clearly stated, in line with the relevant legal and ethical requirements where applicable (e.g., subject consent, permissions, restrictions, etc.).