October 5, 2023

Mixed Methods DMP Example

A Data Management Plan created with DMP Assistant

Data Management Planning Expert Group

Information

Creators: Ceri MacMillan and James Doiron

Affiliation: University of Alberta

Funder: Not Applicable

Template: Portage Template for Mixed Methods (Surveys & Qualitative Research)

Project abstract: This is a fictional data management plan exemplar focusing on mixed methods research (survey & interviews and focus groups), developed using the template for mixed methods research available within the DMP Assistant tool. This exemplar DMP was developed by James Doiron (Research Data Management Strategies Director, University of Alberta Library; Co-chair, Digital Research Alliance of Canada DMP Expert Group) and Ceri MacMillan (University of Alberta Library) for educational and guidance purposes. The premise of the project being described is that the study topic of interest is not sensitive in nature, and that participants have provided informed consent for the de-identified information collected from them to be deposited for long-term preservation, discoverability, and open access.

Identifier: 10307

Last modified: 17-05-2023

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.

There are a range of policies pertaining to the management of research data that are applicable to this project and that are adhered to. These policies are outlined here.

Tri-Agency Research Data Management (RDM) policy: https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management/tri-agency-research-data-management-policy

Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans - TCPS 2 (2022): https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2022.html

University of Alberta Research Records Stewardship Guidance Procedure: https://policiesonline.ualberta.ca/PoliciesProcedures/Procedures/Research-Records-Stewardship-Guidance-Procedure.pdf

We will consult with expertise at the University of Alberta to ensure that we are adhering to all relevant policies and will additionally provide our Research Ethics Application alongside any data deposited for preservation, discovery, and appropriate access and re-use.

We also acknowledge that the University of Alberta has recently implemented an Institutional Research Data Management Strategy, as a roadmap for the future of research data management, the strategy can be found here https://www.ualberta.ca/research-innovation/plans-reports-policies/u-of-a-research-data-management-strategy-document-0501231.pdf

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.

When using acronyms within your DMP (Data Management Plan) it is always recommended to spell them out at least once within any given category so that readers understand what the acronym is referring to.

The data will consist of project data, survey data, semi-structured interviews and focus groups.

Project data includes identity lists of participants, working documents and other project work that will also be collected. Any other potential data not mentioned here that is collected throughout the project will be incorporated into the DMP as it becomes available. The research data collected will produce tabular data, video audio, and text based data (including transcriptions for all interviews and focus groups).

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.

We have searched a wide range of disciplinary-specific, institutional (Borealis Dataverse: https://borealisdata.ca/dataverse/ualberta), and national repositories (Federated Research Data Repository (FRDR): https://www.frdr-dfdr.ca/repo/ and Statistics Canada public use microdata files) for any existing data that may be re-purposed or integrated into our project but were unable to find any relevant data. The project team also consulted with Anna Bombak, Digital Content Specialist within the University of Alberta Library, to confirm that there are currently no existing data that the team can reuse or repurpose.

Is it 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.

Survey data will be collected using the REDCap software platform which is hosted by the Women and Children's Health Research Institute (WCHRI) located at the University of Alberta and is available to University of Alberta researchers. Survey participants will use a secure and unique web-based link to access the survey, and all information they provide will be entered and saved directly into the REDCap system.

REDCap is a browser based electronic data capture software that is widely used for academic research purposes and supports the development and implementation of surveys. For more information regarding WCHRI's REDCap instance see: https://www.wchri.org/redcap

Interviews and focus groups will be conducted virtually using Zoom, a communication platform that enables people to connect virtually using video, audio, phone and chat. Our responses will be recorded directly to a local encrypted computer, following the University of Alberta's Encryption Procedure (https://policiesonline.ualberta.ca/PoliciesProcedures/Procedures/Encryption%20Procedure.pdf#search=encryption).

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.

All transcriptions of the audio visual recordings will be performed internally by project staff using a standardized process. Raw transcriptions will be verbatim copies of the audio visual recordings. De-identification of the raw transcripts will occur during the processing stage of the research project, and using a standardized protocol that will involve removing all direct and indirect identifying information, resulting in the creation of master transcripts that will be used for analytic and dissemination purposes. Additional documentation will accompany the transcription including the name of transcriber, their role within the project and any other relevant demographic information for research purposes. The transcriptions and any associated data will be stored, accessed, and worked within the project's secure virtual research environment, to which only approved project members will have access.

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

> It is considered best practice, and in some instances it is required, to clearly define the length of time that data will be stored on mobile devices (e.g., laptops, electronic pads, digital voice recorders), as well as how they will be transferred to your research space.

The survey data collected using the REDCap software platform will be securely transferred via a secure FTP (File Transfer Protocol) within 48hrs of collection to the project's cloud based virtual research environment (VRE) provided by the Digital Research Alliance of Canada's cloud storage https://alliancecan.ca/en/services/advanced-research-computing/accessing-resources/rapid-access-service

Interview and focus group data collected will be securely transferred within 48 hours of being collected using a secure file transfer protocol (SFTP) platform and will be stored and accessed by project researchers and staff within the project's virtual research space located on the Digital Research Alliance of Canada Rapid Access Service, for more information regarding this service, see: Rapid Access Service: https://alliancecan.ca/en/services/advanced-research-computing/accessing-resources/rapid-access-service

From data collection via zoom, the audio visual files recorded to local computers will be securely transferred via a secure FTP within 48hrs of collection to the project's VRE. All file types will be deleted from the local computer as soon as the file transfer has been completed and verified to ensure no data loss occurred during transfer.

As the data will be transcribed by the research team, there will be no need to move data externally to transcriptionists.

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?

Learn more about the FAIR Principles (Findable, Accessible, Interoperable and Reusable): https://www.go-fair.org/fair-principles/

Our file formats will exist both in non-proprietary and proprietary formats. The non-proprietary formats will ensure that these data chosen for long term preservation and access are available to be used by anyone wishing to do so once they are deposited and made openly available.

Survey data will exist in .csv (non-proprietary), MS Excel, & SPSS (both proprietary) formats. Learn more about SPSS at https://en.wikipedia.org/wiki/SPSS

Any survey data deposited for sharing and long-term access will be in .csv format, promoting FAIR principles allowing the data to be open and shared without requiring proprietary software.

Interview & focus group data will exist in .mp4 (non-proprietary), MS Word & Nvivo (both proprietary) formats. See more information regarding Nvivo from https://en.wikipedia.org/wiki/NVivo

The final de-identified versions of the interviews and focus groups transcripts will be exported into a basic non-proprietary text format (.txt) for deposit, and long-term preservation and access.

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.

Protocols and procedures relating to file and folder structure, naming conventions, and version control will be developed and implemented, and all project staff and researchers will undergo training regarding these processes.

Survey data will be collected within the REDCap program, which has the ability to develop and export a data dictionary that outlines all codes and variables within the survey. Key documentation related to variables will be automatically populated within the survey data, including time and date stamps and other key information to support data management and analytic activities. Upon export from REDCap, survey data will be named as 'raw' data. There may be some additional and yet-to-be-determined versions between the raw and final data. The final and processed data will be named and versioned as 'master' data. From the master survey data, a variety of analytic data files will be created, each adhering to a file naming convention developed to support both management and analysis of data and will be clearly outlined within the aforementioned protocols.

For our qualitative research data (interviews and focus groups) digital audio-visual files will be named and versioned as 'raw audio visual', while the verbatim transcripts produced from these will be named and versioned as 'raw transcript audio visual'. Once a raw transcript has been processed, with any identifying information removed, it will be named and versioned as a 'master transcript'. Copies of the master transcripts will be created and used for analysis purposes.

All qualitative interviews will include summary information including: name of data collector, location of interview, and the date and time that the interview was conducted. Additionally, qualitative interviews will have accompanying notes containing key contextual information and metadata. Documentation explaining naming conventions, metadata standards used, the software used throughout the project, and any processes for accessing these data will also be created and updated as needed throughout the project and a final anonymized copy will accompany any data destined for long-term preservation.

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.

Having clear file naming and versioning protocols will help to ensure that your data are well organized, that participant confidentiality is maintained, and support all research activities, including data collection, data processing, analysis, and deposit.

File naming documentation will be developed and implemented. Components of file names will include, as needed: file version (raw, edit, master, analytic), date (ie., dd/mm/yyyy), and any applicable key contextual information, (e.g., geographical location, interviewer initials or code) and final master copies will include in the name whether the data contained within have been de-identified. If there are any changes, a version number will also be included within the file name. The file names will be in all capitals and the format: NAMEOFVERSION_DATE(DD/MM/YYYY)_CONTEXTINFORMATION_VERSIONNUMBER.file extension

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

We will develop and implement clear metadata documentation protocols prior to starting data collection which will clearly communicate expectations, standards, and processes for capturing and implementing documentation to support the research project. Some example topics that these protocols will cover include file naming conventions, file versioning, folder structure, and both descriptive and structural metadata.

Researchers and project staff will have the opportunity to contribute to these protocols, and once they are finalized the protocols will be implemented to support the active phases of the research project. Documentation will be kept in clearly named folders within the virtual research project folder space and will be accessible by all researchers and staff.

One member of staff, whether that will be with the acquisition of a specific data management personnel, or someone already involved in the project will be the lead for ensuring the documentation and metadata standards used are consistently updated according to policies and procedures.

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

> Don't worry if you don't have all the answers yet. The DMP is a living document and is expected to be updated as new developments occur during the lifespan of the research project.

After consulting with the metadata team located within the University of Alberta Library, we concluded that the Data Documentation Initiative (DDI) metadata standard would optimally support our survey data, especially as our survey data collection platform, REDCap, also employs DDI metadata standards. More information on DDI can be found at: https://ddialliance.org

It is not yet determined what metadata standards will be used for supporting qualitative research. We will be consulting with on-campus expertise within the University of Alberta Library to determine if there are any other standards that will be beneficial for the project.

Storage, Access, and Backup

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.

All data storage and backup procedures will be clearly outlined within the project's data collection policies and procedures which will be developed prior to data collection. These procedures will also indicate where data will be stored throughout the active stages of the project.

As we are collecting survey data using REDCap software, the raw data will be transferred using a FTP process, and will be stored securely. REDCap servers are locally hosted by WCHRI within the Faculty of Medicine and Dentistry and undergo regular backups (incremental and full). Our virtual research project space has a regularly established scheduled incremental and full backup process in place to ensure no data loss occurs.

Qualitative interviews will be conducted using virtual recordings over Zoom. Upon completion of the interviews the data will be securely transferred within 48 hours to the virtual research project space located on Digital Research Alliance of Canada's cloud platform. Once the interviews are uploaded to the cloud platform they will be permanently deleted from the local computers on which the interviews were saved. This VRE undergoes backups on a regular schedule which include incremental and full backup processes.

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

All data will be securely stored on the Digital Research Alliance of Canada's (DRAC) cloud platform and these data will be accessible only by approved researchers, trainees, and project staff. Access to the platform is securely password protected, with access rights ultimately approved by the Principal Investigator and assigned by DRAC. Researchers will be able to access this VRE remotely, and any software required to complete the research project is contained within this VRE, so no additional software will be needed.

For more information regarding this service provided by Digital Research Alliance of Canada see:
https://alliancecan.ca/en/services/advanced-research-computing/accessing-resources/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.

> Estimating as early as possible how much data you will be collecting is important as this can help you to have a clear sense of how much storage space you require both during the active phase of your research project, as well as to support the long-term preservation of your data.

We estimate that we will be collecting approximately 200 surveys, 30 interviews (approximately 30 min in length each), and 3 focus groups (approximately 90 min in length each). Total magnitude of data, including accounting for versions (raw, master, analytic) is estimated to be under 20 GB. Other documentation, including transcriptions and project management documents, accompanying the data is estimated to require around 5GB of storage. Total storage requirements: ~25GB.

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.

All data will be maintained for a minimum 5 years after the completion of the project, as per the University of Alberta ethics requirement policy: https://policiesonline.ualberta.ca/PoliciesProcedures/Procedures/Research-Records-Stewardship-Guidance-Procedure.pdf

The surveys collected will be anonymous, with no direct or indirect identifiers present. Survey data will undergo data quality checks, including missing and out of range values, logic rule and skip pattern violations, and derivation of scales. When converting survey data between formats there will be systematic checks - both case and variable wise - to ensure that no data is lost.

The video interviews will be transcribed by research assistants - this will be a verbatim transcription, with no contextual information removed. However, interview participants will be de-identified by referring to them by predetermined codes or pseudonyms. There will be no video component to the video files saved as participants will not be required to turn on cameras, therefore only the audio visual will be saved. Any other information that could potentially identify participants shall also be removed. Once the initial verbatim transcription is completed it will be saved and versioned as a 'raw' transcript. This raw transcript will then be provided to the interviewer to review for completeness and to remove any necessary indirect identifiers from the text that may have been missed. Once this has occurred it will be named and versioned as a 'master' transcript. Information about the transcription process, including the names of the transcribers, will be stored and available upon request from the research team.

For long term storage and preservation, we will be preserving both the raw and the master (cleaned, processed and de-identified) versions of the surveys. We will be preserving only the approved and de-identified interview and focus group transcripts (not the audio or video files), and these will also be deposited for long-term preservation and open access.

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, including those details.

We will be using the University of Alberta's instance of Borealis, the Canadian Dataverse Repository, (https://borealisdata.ca/dataverse/ualberta) to deposit and support the long-term preservation, discovery, and access of our data. This Dataverse is freely available for use by our project, and contains a number of desirable features including the assignment of unique and persistent digital object identifiers (DOIs), the ability to restrict access to data at the file level,

built in data citations, data usage metrics, and file versioning. More information regarding Dataverse is available at: https://en.wikipedia.org/wiki/Dataverse

Should the U of A instance of Dataverse be determined to not be able to support any of our data deposit needs, we will be consulting with the U of A Library data team to help guide us through the deposit process in order to find any optimal solutions and supports for our project.

Sharing and Reuse

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

We plan to share all processed survey data. This will include responses at the participant level, with all direct and indirect identifiers removed, and will include derived variables used for analytic purposes. The de-identified and processed versions of surveys will be deposited for long-term preservation and open access in our instance of Dataverse. We will be preserving only the de-identified interview and focus group transcripts (not the audio visual files), and these will also be deposited for long-term preservation and open access.

We will obtain participant consent in order to share the data, and will additionally be depositing and making available an example of both our study information letter and participant consent form, as well our approved institutional ethics application.

The data available for sharing will be in non-proprietary formats to promote accessibility and re-use of the data in accordance with the FAIR principles.

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.

All de-identified data will be openly discoverable and accessible to anyone via the University of Alberta Dataverse repository platform hosted by Borealis, the Canadian Dataverse Repository. If any restrictions are requested by participants, we will try to accommodate these restrictions, and include any restrictions into the project DMP.

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

All data will fall under open data licensing (CC-BY 4.0), unless any changes occur during the data collection stage of the research project which necessitate changing the type of end-user license.

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.

It is important to think ahead and identify who will be involved in the management of your data. Not all projects are the same but, to varying degrees, all may require dedicated resources to help effectively manage the data!

Research data management will be a shared responsibility and will involve the Principal Investigator (PI), co-investigators, collaborators, trainees, and research project staff. Additionally, we plan to have a dedicated research data manager position (estimating 0.6 FTE). We will be including this position into our funding application and budget. This position will assist in all research data management processes and will report to the PI in order to ensure all data is managed appropriately. It is the responsibility of everyone working on the project to be aware of, and practice good measures in relation to research data management.

We will be creating a Research Data Management Committee (RDMC) that will be responsible for data governance, including developing policies and procedures relating to research data management and that will be implemented throughout the project. We plan to conduct a 1-1/2 day research data management training event prior to any data being collected, and all investigators, trainees, and research staff will be required to participate in this training.

All research staff involved in collecting data (both survey and qualitative interviews/focus groups) will sign confidentiality agreements. Analysts, investigators, and trainees that will be handling data for analysis and dissemination purposes will additionally undergo training and signing of confidentiality agreements.

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?

Should personnel changes occur we will refer to our Research Data Management Committee (RDMC) as well as the RDM Policies & Procedures governance which help to identify both primary and secondary individual roles and responsibilities for various RDM activities. In the event that new team members/research staff/trainees join the project, we will similarly refer to these policies and procedures in providing direction for supporting onboarding. All new members of the research project will undergo in-depth RDM training and signing of confidentiality agreements.

Should the PI leave the institution prior to the conclusion of the project, she/he/they will work closely with the RDM Committee, as well as institutional administrative research support, to identify how to best support the research and management of the project data moving forward, while also adhering to all ethical, legal, and contractual obligations.

Data stewardship, including preservation and sharing after the project has concluded, is a shared responsibility, though ultimately the Principal Investigator is responsible for ensuring that these data move forward with preservation and are appropriately shared. This will occur with the support of the RDM Committee, as well as collaboratively with the University of Alberta Library.

Data will be deposited into the U of A Library instance of the institutional research data repository, Borealis, the Canadian Dataverse repository. All data deposited will be open access and U of A Library will effectively become the stewards of these data, ensuring their continued preservation and access.

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

This is simply an example demonstrating where a project has deemed that a dedicated research data manager position is necessary. Not all projects will need, or can justify, having a dedicated position such as this. It is still important though to identify a project's RDM needs and from there determine what supports may be required as well as to identify who among the research team will be responsible for specific activities (e.g., investigators, research staff, graduate students).

The resource needs that we have currently identified are:

- A dedicated research data manager (~0.6 FTE) for a period of 12 months. We will be including this position into our funding application and budget at an overall amount of $55,000.

- Cloud based virtual research environment space. This will be provided by Digital Research Alliance of Canada (DRAC) via their Rapid Access Service (RAS). https://alliancecan.ca/en/services/advanced-research-computing/accessing-resources/rapid-access-service

- Proprietary analytic software both for the survey (SPSS) and qualitative data (Nvivo). We are not yet sure if we will need to pay for license costs for these, and so will be consulting with DRAC on this. If so, we are estimating that we will require 3 SPSS licenses and 2 Nvivo licenses for a period of 1 year and will need to determine the cost of these as we would require server based licenses. We will determine these costs and include them within our funding application budget.

- Zoom accounts can be utilized for interviews and focus groups without incurring any additional costs. Group members can use the laptops for the project that have an associated cost of around $1500.

- Borealis, The Canadian Dataverse repository, is a free to users platform that we will be using for long-term preservation and storage once the project is complete.

- We will update this DMP as needed if any components of the plan change, for this we used the free DMP Assistant tool. See https://assistant.portagenetwork.ca for more information.

Ethics and Legal Compliance

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

All of our research data will have any potentially identifying material removed, and both raw and master copies will be available in open access format on Borealis, the Canadian Dataverse repository platform. As data is subject to a CC by 4.0 license, it can be used and adapted as needed for secondary usage. The question of secondary usage will be identified on our participant consent forms and will determine any restrictions on use placed on the data.

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

In addition to following the legal and ethical policies set out by the Research Ethics Board at the University of Alberta, our data will remain stored within Canada, and is therefore protected by Canadian laws. All data chosen for long term preservation will be open access but will have any potentially identifying information removed to protect the research project participants. Data is subject to CC by 4.0 license, and will be dependent on participant consent for data re-use.