

DMP title

Project Name CDCF Workshop

Project Identifier 8675309

Grant Title 12345

Principal Investigator / Researcher James Doiron

Project Data Contact James Doiron

Description This is an example data plan used for educational purposes at the Canadian Data Curation Forum held in Hamilton, Ontario on October 16, 2019.

Institution University of Alberta

Data Collection

What types of data will you collect, create, acquire and/or record?

We will be collecting surveys which will then be exported into tabular format.

We will also be conducting interviews which will then have both audio and text (transcriptions) data associated.

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

Our file formats will be:

- Surveys (.csv, MS Excel, & SPSS)
- Interviews (.mp3, MS Word, Nvivo)

The raw versions of the survey are in .csv format and so are non-proprietary and will be deposited for long-term preservation and access.

The raw audio files are in .mp3 (non-proprietary) format. The final de-identified versions of the transcripts will be exported into a basic non-proprietary text format for deposit, long-term preservation and access.

If data are collected using laptops or mobile devices, please explain how will you will securely store and transfer the data.

All laptops are securely encrypted and will only be used for data collection, and **not for storage**. Data collected will be securely transferred within 48 hours of being collected using either an encrypted hard drive/key or a secure file transfer protocol (SFTP) site. Data will then be stored on secure servers on [Compute Canada's cloud Rapid Access Services](#) system where access will only be available to researchers, trainees and research staff.

How much data do you anticipate collecting? Include an estimate of how much storage space you will require (in megabytes, gigabytes, terabytes). This estimate should also take into account storage space required for file versioning, backups, and the growth rate over time.

We estimate that we will be collecting 800 surveys and 20 interviews (approximately

30 min in length each. Total magnitude of data is estimated to be under 10GB.

Are there any existing data that you can re-use? If so, please explain how you will obtain that data and integrate it into your research project.

We have searched for existing data that may be able to be re-purposed but for our project but were unable to find any. We searched a wide range of disciplinary, institutional (Dataverse) and national repositories (FRDR, Stats Canada DLI, ICPSR, etc).

P.S. - We also contacted the data team located within the UofA libraries. James talked to himself there.

What conventions and procedures will you use to structure, name and version control your files to ensure that your data is well-organized?

We are not yet sure what we will be using for metadata standards, document naming, version control, etc.

Documentation and Metadata

What documentation will be needed for the data to be read and interpreted correctly in the future? This includes study-level documentation, data-level description, and any other contextual information required to make the data usable by other researchers

Survey data will be collected within [REDCap](#), a secure data collection and management software hosted by the Women & Children's Health Research Institute (WCHRI) at the University of Alberta. REDCap features include the ability to develop and export a data dictionary which will outline all codes and variables within the survey.

All qualitative interviews will include summary information including: data collector, location of interview, and the date that the interview was conducted. Additionally, qualitative interviews will have accompanying field notes containing contextual information and metadata.

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), location (EDM, HAM, CAL), and interviewer (initials or code).

Please list the metadata standard and tools you will use to document and describe your data. If there is not an appropriate standard, please explain how you will ensure consistency in your documentation.

[The Data Documentation Initiative](#) (DDI) metadata standard will be used for survey data.

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 Libraries, as well as the [Qualitative Data Repository](#) (QDR) located at the [Maxwell School of Citizenship and Public Affairs](#) at [Syracuse University](#)

How will you make sure that documentation is created or captured consistently throughout your project?

We will develop and implement clear metadata documentation protocols prior to starting data collection. Researchers and staff will have the opportunity to contribute to these, and once they are finalized they will be implemented to support the active phases of the research project. Documentation will be kept in a clearly named folder within the research project folder space and will be accessible by all researchers and staff.

Storage and Backup

How will your data be stored and backed up during your research project?

All data storage and back-up procedures will be clearly outlined within the project's data collection policies and procedures which will be developed prior to data collection.

Survey data will be collected using [REDCap](#), a secure data collection and management software hosted by the Women & Children's Health Research Institute (WCHRI) at the University of Alberta. REDCap undergoes regular (daily, weekly, monthly) backups, and supports secure transferring of data using both encryption and a secure File Transfer Protocol platform. Upon being exported from REDCap, survey data will be immediately ingested into our research project space located on Compute Canada's cloud platform.

Qualitative interviews will be conducted using encrypted digital voice recorders. Upon completion of interviews they will be securely transferred within 48 hours to the research project space located on Compute Canada's cloud platform via direct upload. Once the audio interviews are uploaded to the cloud platform they will be deleted from the digital voice recorders.

How will you ensure that sensitive data is stored securely and only accessible to the research team during the research project?

All data will be securely stored on Compute Canada's cloud platform which will be accessible only by researchers and study staff. Access to the platform is securely password protected, with access rights ultimately approved by the Principal Investigator and assigned by Compute Canada.

Preservation

Which data are selected for preservation and access will depend on potential reuse value, whether there are obligations to either retain or destroy data, and the resources required to properly curate the data and ensure that it remains usable in the future. In some circumstances, it may be feasible to preserve all versions of the data (e.g. raw, processed, analyzed, final), but in others, it may be preferable to only keep only selected data (e.g. transcripts instead of audio interviews).

All data will be maintained for a minimum 5 years after the completion of the project, as

per UofA ethics.

For the long term, we will be preserving both the raw and the master (cleaned and processed) versions of the surveys. The de-identified and processed versions of surveys will be deposited for long-term preservation and open access. We will be preserving only the de-identified interview transcripts, and these too will be deposited for long-term preservation and open access.

At the end of your research project, where will you deposit your data for long-term preservation and access?

At this time we believe that we will be using the University of Alberta's institutional data repository, [Dataverse](#), to deposit our data. Dataverse is freely available for use by our project, and contains a number of desirable features including the assignment of digital object identifiers, ability to restrict access to data, built in data citations, data usage metrics, and file versioning.

Should the UofA Dataverse be determined to be able to support any of our data deposit needs, we will be consulting with the UAL data team to help guide us through the deposit process in order to find the optimal solutions and supports for our project.

Please describe how you will prepare the data for preservation and access, including any necessary procedures for data cleaning, normalization or de-identification. Explain how you will prevent data from being lost while processing and converting files.

The surveys collected are anonymous, with no direct 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 are lost.

The audio 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 as P1. Once the initial transcription is completed it will be a raw transcript. This will then be provided to the interviewer to review for completeness and to remove any necessary indirect identifiers from the text. Once this has occurred it will be a master transcript.

Data Sharing and Reuse

What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final). Consider which data may need to be shared in order to meet institutional or funding requirements, and which data may be restricted because of confidentiality/privacy i

We plan to share processed survey data. This will include responses at the participant level, with all direct identifiers removed, and including derived variables using for analytic purposes.

Regarding the qualitative interviews, we plan to share de-identified transcripts.

We have received participant consent for sharing these data, and will depositing and making available an example of both our study information letter and participant consent form.

How will you be sharing your data? (e.g. institutional repository, a specialized data archive, project website, informal/on-request sharing). Include a brief description of any resources needed to share your data (equipment, systems, expertise, etc.).

We plan to share our data using the University of Alberta institutional data repository, Dataverse. Dataverse is freely available for use by our project, and contains a number of desirable features including the assignment of digital object identifiers, ability to restrict access to data, built in data citations, data usage metrics, and file versioning.

Should Dataverse be determined to be able to support any of our data deposit needs, we will be consulting with the UAL data team to help guide us through the deposit process in order to find the optimal solutions and supports for our project.

Please 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 data will be openly discoverable and accessible via the University of Alberta Dataverse repository platform.

What type of end-user license will you include with your data? Please include a copy of this license with your Data Management Plan.

All data will fall under open data licensing (CC-BY)

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 will be shared responsibility and will involve the PI, co-investigators, trainees, and research staff. We will be creating a Research Data Management Committee (RDMC) that will be responsible for 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) will sign a confidentiality agreement. Analysts, investigators, and trainees that will be handling data for analysis and dissemination purposes will additionally undergo training and signing of confidentiality agreements.

What will happen when personnel changes occur or if the principal investigator leaves the institution before the project has concluded?

Should personnel changes occur we will refer to our RDM Policies & Procedures governance which will inform us of optimal coverage by identifying both primary and secondary individual roles and responsibilities for various RDM activities. In the event that new team members/research staff/trainees come on board, we will similarly refer to these policies and procedures in providing direction for supporting optimal transitions. All new members of the research project will undergo indepth RDM training and signing of confedetraility agreements.

Should the PI leave the institution prior to the conclusion of the project, the existing PI will work closely with the RDM Committee, as well as his own institutional support, to identify the optimal scenario for supporting the research and management of its data moving forward. All RDM policies and procedures, including plans for the deposit and long-term preservation of research data will remain intact.

Who will be responsible for data sharing and preservation after the project has concluded? Indicate the party who will have primary responsibility for how the data will persist over time when the original personnel have moved on.

Data will be deposited into the UofA Libraries' Dataverse, and/or another trusted digital research data repository.

What resources will you require to implement your plan? Will extra people, time or hardware, storage be required? How much will this cost (estimation)?

Question not answered.

This document was generated by DMP Assistant (<https://assistant.portagenetwork.ca>)