Data Management Plan: FoM Basic REDCap Research project

This example Data Management Plan (DMP) was created by the Faculty of Medicine Digital Solutions Data Management (FoMDS DM) Team at the University of British Columbia, with the purpose of providing a standardized DMP model for a research project using FoM REDCap. Here, fundamental processes and procedures core to the health research have been incorporated as a starting point for the faculty/research teams when developing their research projects' DMP's. This example demonstrates the utility of a model DMP being used by a research group or collective to maintain best practices in data management.

This DMP was created using the questions from DMP Assistant (Portage). For more information visit UBC Research Data Management DMP Assistant

Project Details

Project Name: Insert full Project Title

Project Abstract: Insert Project Abstract

Example: The **XX Research team** *at the University of British Columbia carries out fundamental research in [describe research] to support XXX. [insert research team] research activities cover [describe high level research objective].*

Research Domain: *Medicine and Health Sciences (or whichever options suits your project best)*

Project Start: *Enter date*

Project End: *Enter date*

Identifier: *A pertinent ID as determined by a funder and/or organization*

Funder: *If applicable, enter Funding body. For example: UBC Strategic Investment Fund*

Funding Status: *if funded, select status*

Grant Number/URL: *Clinical trial number or link to trial. For example: clinicalttrials.gov/study/NCT03XXX*

Contributors

Add all project Contributors: include names, email, Orchid ID. Affiliation will prepopulate as UBC; therefore, update for external contributors

Consider the following project roles

- **Principal Investigators**
- **Data Manager**
- **Project Administrators**

- **Any other team members**

If you are using acronyms, definitions can be defined in a simple table

Acronyms used in this document			
CRF's	Case Report Forms	**SOPs**	Standard Operating Procedures
.CSV	Comma Separated Value File	**.XLS**	Microsoft Excel Spreadsheets
FoM	Faculty of Medicine	**REDCap**	Research Electronic Data Capture
HR	Heart Rate	**PI**	Principal Investigator
DM	Data Manager	**RA**	Research assistant
REB	Research Ethics Board	**UILO**	University-Industry Liason Office

Data Collection

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

1. Participant metadata
    - Content from the consent form, including first and last name, date of birth and contact information such as email address
2. Demographic data, such as education, socioeconomics and gender identity
3. Data from self-completed questionnaires, including detailed indicators of medical history
4. Data from in-person health assessments, including:
    - HR data

A complete itemized list of the variables collected/generated for this study is tracked in a study codebook

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?

Non-proprietary or open format data will be collected. Data is gathered on paper and other data from the patients' charts. Other data is collected directly from patients via surveys/CFRs and from medical devices. All data collected will be entered into REDCap; therefore, saved as open format.

What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

Folder naming conventions

Folders are named according to data type.

- Example: HR data

File naming conventions

● REDCap exports are saved using the following naming convention: Studytitle_DATA, year-month-day, time, file extension.

    ● Example: REDCapDemo_DATA_2022-10-11_1344.CSV

Documentation and Metadata

What documentation will be needed for the data to be read and interpreted correctly in the future?

- The content and functionality of the REDCap database are documented in the data dictionary and study codebook. The majority of data items, including data from in-person health assessments, self-completed questionnaires, and clinical data from charts, are referenced and stored using the indexing and naming conventions described by the project's REDCap data dictionary. The REDCap data dictionary is a .CSV file that represents the detailed structure of the database and is used to program its functionality. The REDCap study codebook is a human-readable data dictionary version that allows quick reference to the variables and functionality detailed by the data dictionary.
- Notes capturing miscellaneous documentation of special cases in the data are recorded on CRFs to participant files and comments, which are subsequently transcribed into REDCap.
- REDCap will be maintained in accordance with the Standard Operating Procedures (SOPs) developed by the UBC FoM Data Management Team

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

Methodologies for data collection/generation

Data collected during in-person visits are conducted by appropriate trained and delegated Research Assistants (RAs).

Quality assurance

SOPs are developed for data collection steps that are to be followed by research team. These SOPs are stored in the online repository that all research team can access and review at any time. Files are updated as necessary, and version controlled.

Data Quality assurance checks will be conducted by research team.

- Manually entered participant data is checked again the original data collection source for accuracy.

- Data quality rules are built in the project and ran on a regular cadence to ensure the entered data meets the variable requirements such as accuracy, completeness, consistently and validity.
- Random consent forms and CRFs will be checked for accuracy and completeness
- RA's conducting data entry into REDCap will report any irregularities they find to the Data Manager (DM), during the completion of their data entry duties. The DM will issue query reports to the RAs, who will work to resolve identified irregularities and then return the reports to the DM, who will retain them for documentation purposes.

If you are using a metadata standard and/or tools to document and describe your data, please list here.

Documentation and data description processes follow in-house standards, as outlined and maintained in the REDCap data dictionary and study codebook. These documents can be shared with any data made available for subsequent use to ensure readability and operability.

Data will be standardized whenever possible, shared data will be deidentified, and original data will be maintained at the University of British Columbia. Common Data Elements will be applied when applicable, using the Clinical Data Interchange Standards Consortium Data (CDISC) standards.

Borealis Dataverse metadata schema (DDI) will be applied once study is archived.

Storage and Backup

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

Data storage

All digital data will be stored in UBC FoM REDCap and is managed by the FoMDS Data Management Team. The FoM REDCap servers are located at the UBC University Data Centre (UDC) and use SSD (Solid state drives) for storage.

The PI has locked laboratories at the University of British Columbia Point Grey campus at the Research Center which will house all paper documents.

Backup

The digital data will be stored in FoM REDCap and backed up according to UBC Data Backup and Retention Policy. Backups of the data are taken at regular intervals and held on multiple devices, with copies held in different secure locations.

What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

For this project, the estimated data storage is 25MB. This is for the digital data for one complete copy of the dataset. The digital data volume is well withing the existing storage capacity of the research group.

Similarly, the paper documents generated (consent and CRF's that are not completely electronically) are expected to fill one locked filing cabinet, owned by the PI. However, if this is underestimated, all digital and paper data storage locations can be scaled to fit within the current capacity without difficulty.

Paper documents will be retained and the PI's storage facility at the University of British Columbia Point Grey campus for 10 years after data collection is completed.

Digital data will be kept on the PI's allocation of the University of British Columbia and Research Centre XX data servers, and the backup devices for the study, for 10 years after data collection is completed. Digital data stored in FoM REDCap will be stored for a minimum of 10 years after the project is archived in REDCap.

How will the research team and other collaborators access, modify, and contribute data throughout the project?

Study processes and protocols are in place for accessing each type of data collected, in conjunction with guidelines for which project roles have responsibility for data modification and contribution.

Data will be entered online and directly in FOM REDCap via a user account. User accounts are secure requiring a username, password and Multi Factor Authentication (MFA). User accounts will be created directly by the research team administrator following the FOMDS SOP: Enabling REDCap User Access. Each user will be given a role or a collection of user rights with specific access to the Faculty of Medicine REDCap project. Continued access will be controlled by the research team administrator, such as the Data Manager.

Individual team members and collaborators will not store data on their own devices, in physical spaces, or outside of the project REDCap space.

Data Preservation

Note: UBC Library offers Data Repository options: https://researchdata.library.ubc.ca/deposit/

All de-identified demographic and clinical data could be shared via the institutional repository UBC Dataverse which provides management of metadata, persistent identifiers (DOIs), and long-term access. UBC Dataverse is a digital repository for presenting and preserving the scholarly work and is administered and maintained by the UBC Library. All Borealis files are digitally preserved with Level 1 digital preservation.

Where will you deposit your data for long-term preservation and access at the end of your research project?

Paper documents will be retained at the PI's storage facilities at the University of British Columbia Point Grey Campus for 10 years after data collection is completed. After this, the paper will be confidentially shredded.

Digital Data will be kept on the PI's allocation of the University of British Columbia's and affiliated research Center's data servers and the back-up devices for the study for 10 years after data collection is completed. After this the data will be deleted and the storage devices will be reformatted.

At the end of the study, the project will be deposited for long-term preservation into the University of British Columbia's Dataverse to make the data available to collaborators. Dataverse is a university-administered and controlled access repository that facilitates research data creation, management, and dissemination. The University of British Columbia's version of Dataverse is jointly hosted in Canada in three geographical locations ensuring data integrity.

Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, and inclusion of supporting documentation.

Where possible, all data will be stored in open or industry standard file types (i.e. .CSV or .XLS). The original data collected will be retained in its original pre-processes format (i.e. .CSV) utilizing the previously described naming conventions for anonymization. Any processed files will contain information denoting processing procedures, decision and the date of the analysis or processing.

- Data kept in REDCap database are exported as .CSV text files, which is an open format that is readable in Microsoft Excel
- Heart rate data are exported from Garmin Heart Rate Monitors and save in .CSV formats

Sharing and Reuse

What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).

Raw, processed and analyzed datasets may be shared, in compliance with ethics documentation. Analyzed data will be shared in UBC Dataverse. Data sharing will comply with end-user licenses developed for the study; see the following answer.

Have you considered what type of end-user license to include with your data?

Before research data is shared, the PI or research team should work with the University of British Columbia University-Industry Liaison Office (UILO) to request their involvement in creating or using a template Data Transfer Agreement. The collaborating parties and their respective institutions must sign and adhere to these agreements. The University of British Columbia Research Data Management Services may also be consulted during the drafting of agreements to ensure that the wording is consistent with the policies that govern the usage of Dataverse.

What steps will be taken to help the research community know that your data exists?

Study information and discoveries will be made available to potential collaborators through public postings, conferences, and presentations. Publications will describe the goals, methods, and planned analyses of the study and include the contact information of the PI. A posting for the study on the clinicaltrials.gov database already exists (ClinicalTrials.gov Identifier: NCT0393XXXX), which informs researchers that they can request access to project data. A protocol paper describing the trial published in XX Journal (DOI: xxx) also informs researchers that they can request access to our data. Once analyzed data has been deposited in Dataverse, DOIs will allow for the data to be searched as well as linked to publications.

Responsibilities and Resources

Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.

The PI assumes ultimate legal and ethical responsibility for all aspects of data management. The PI's responsibilities include developing and implementing data collection and management procedures that ensure the data are credible and that the rights and safety of participants are preserved.

The PI may delegate staff to manage aspect of the study's data management on his behalf, except for the assumption of ultimate legal and ethical responsibility. The PI or delegates may be responsible for day-to-day oversight of data management activities and data sharing, ensure the timeliness of data entry and review data to ensure the quality of data entry. They may ensure that metadata are sufficient and appropriate and that the data management and transfer plan follow the Good Clinical Practice (GCP) data principles. The PI or delegates are responsible for implementing the data management practices and request approval for a revised plan if there is any deviation from the approved DMP plan.

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?

The study PI will identify a colleague or collaborator with equivalent qualifications to take responsibility for the project in the event they are not capable. In the event the PI is unable to fulfill their responsibilities, the research team will inform the University of British Columbia's Research Ethics Board (REB) of a change in study leadership. The research team will seek guidance from the Board on the appropriate action to take based on the circumstances of the situation and formally request the transitions of responsibilities to the appointed steward.

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

The data management plan will require support from the PI and research team. A full-time data coordinator will be necessary to support the management of research data. This type of Research Data

Management support is built into the grant applications and will cost $60,000/year for the project. Additional support for the FoMDS DM, UBC Research Data Management and Advanced Research Computing will also be provided throughout the course of the project for no charge.

Ethics and Legal Compliance

If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

The main data security risk is the potential for participants' confidential personal health information to be unintentionally disclosed to third parties. This risk is mitigated by storing physical data in secured buildings in locked cabinet and storing digital data on password and encrypted databases (see REDCap details above). During the consent process, participants are made aware of this risk and that despite these precautions the risk still exists; therefore, they can choose to withdraw consent.

All study staff will protect participant confidentiality as part of their employment and complete the necessary UBC Privacy Matters training. They will also follow UBC Data Management and study specific SOPs along with Cyber Security policies.

If datasets are considered sensitive, then data management will be the responsibility of the lead researcher. Data will be subjected to restrictions according to the best practices and protocols outlined by the University of British Columbia Office of Research Ethics . Only appropriate and approved metadata will be made available. Authorized personnel with sensitive data will be encouraged to ensure that files are encrypted.

If applicable, what strategies will you undertake to address secondary uses of sensitive data?

Study participants were provided with an opportunity to consent that their data could be shared with collaborators. For those participants who provided this consent, their data will be made available to collaborators following the creation of Data Sharing Transfer Agreement.

Active research data, as well as sensitive data, may be subject to restriction according to the best practices and protocols outlined by the UBC University-Industry Liaison Office, the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) and/or OCAP principles. The availability of data collected from external collaborators will be subject to Data Transfer agreements made with these collaborators.

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

Any legal, ethical and intellectual properties issues will be guided by relevant policies to the University of British Columbia and the Laws of Canada.

The collaboration parties and respective institutions must sign and adhere to the Data Transfer Agreement. Templates for these can be found on the University of British Columbia University-Industry Liaison Office (UILO) website (https://uilo.ubc.ca/researchers/forms-and-templates). If these templates do not meet the needs of the PI or collaborators, the UILO will be consulted to create a non-templated contract.