

Researcher Data Management Planning Guidance

Table of Contents

1. Purpose	1
2. Data Management Plans (DMPs) and Data Deposit	1
3. General Data Management Plan Guidelines	1
Appendix 1: Quick Reference for College of the Rockies Student and Staff Researchers: Protecting Electronically Stored Personally Identifiable Research Data	3
Appendix 2: Portage Data Management Plan Exemplar: Ecohydrology Research Group	5
Appendix 3: Portage Data Management Plan Guidance: Qualitative Mixed Methods	11

1. Purpose

The purpose of this document is to guide individual researchers and research teams’ research data management plans for research projects. This guidance underlines the importance of data as an output of research.

A research data management system is the complete system under which research data is collected and managed and includes the formal processes of institutional assessment and periodic re-assessment, creating, reviewing, and updating procedures, support services, support personnel, longevity of the approach and compliance.

2. Data Management Plans (DMPs) and Data Deposit

Central to the research data management practice are individual research data management plans (DMPs).

A DMP is a formal document that details the strategies and tools researchers will implement to effectively manage data during the active phase of research, and the mechanisms they will use for preserving and appropriately sharing data at the end of the project. A DMP is a “living” document that can be modified throughout the project to reflect any changes that have occurred.

College researchers will use formal, standardized data management plans as part of the research process. At the beginning of a research project, the Principal Investigator should ensure the following is completed prior to starting the research.

Guidelines and explanations for data management plan elements are provided below, and examples are given in the appendices.

3. General Data Management Plan Guidelines

- Begin by giving a description of your research project, its focus, and purpose.
- Avoid the extensive use of discipline specific jargon - your DMP should be easily understood by anyone.
- Provide clarification for any acronyms used.
- Provide rationale for decisions made.
- Your DMP is a living document - update it as needed!

a. Data Collection

- Describe the data that you will be collecting, including the type, format, and volume.
- Describe how you will be collecting your data.
- Establish standards for naming and organizing data files, folders, and versions.

- b. Documentation and Metadata (Description of Data)
 - Describe how you will ensure that your data are understandable, interpretable, and usable both by current and future researchers.
 - Provide descriptive information for your data to be discoverable once deposited at the end of your project.
- c. Storage and Backup
 - Describe storage methods and backup procedures for the expected data volume to ensure data security and integrity, and to avoid data loss during the active phase of your research. Refer to Appendix 1: Quick reference for College of the Rockies student and staff researchers: Protecting electronically stored personally identifiable research data.
 - Describe how collaborators will access and work with the data.
- d. Preservation for Long Term Access
 - Decide which data to keep and for how long.
 - Address any obligations to retain or destroy certain data.
 - Describe the resources required to preserve the data for long-term access.
- e. Sharing and Reuse
 - Describe where and how data will be shared (if permissible).
 - Decide in what form (raw, processed, analyzed, final) data will be shared.
 - Select options for sharing agreements or data licenses for reuse.
- f. Responsibilities and Resources
 - Identify who will be responsible for each data management task.
 - Consider resources and costs required both during and after the project.
- g. Ethical and Legal Compliance
 - State how you will store, share, and preserve data in a way that ensures ethical and legal requirements are met.

Appendix 1: Quick Reference for College of the Rockies Student and Staff Researchers: Protecting Electronically Stored Personally Identifiable Research Data

Research participants trust researchers to manage their personal data in a way that is secure and ensures privacy, especially for personally identifiable information. The TCHPS2 (Tri-Council Policy Statement 2018) states “Researchers shall provide details to the REB regarding their proposed measures for safeguarding information, for the full life cycle of information: its collection, use, dissemination, retention and/or disposal”. It is the ethical responsibility of researchers to take appropriate steps to protect these data.

General Guidelines:

1. Decide where to keep different types of participant research information

It is important to keep personally identifiable participant information, or codes that link participants to their data, separate from the actual data. The main reason is that if one device/drive (e.g. a laptop) is accessed, participants cannot be identified by simply matching up the two documents. It is preferable devices be physically stored separately as well. When deciding where to store codes and/or participant data, researchers should know about tools the College offers, along with some important considerations about which ones to use:

- College of the Rockies B: Drive – The B drive is available to faculty, and students with permission by IT services, and is recommended for storing sensitive research data, including any personally identifiable data. Files and folders can be granted restricted access. Data from the B: drive is stored on secure College servers, and can be accessed when connected to the College’s VPN.
- One Drive – Research data may be stored on One Drive as long as it is not personal health information
- Personal Computer or Laptop – If storing research data on a computer, the documents should be encrypted and the computer must be password protected. It would be even better to encrypt the computer. Set a time out to automatically lock after a set number of minutes.

2. Prevent data theft/loss

Theft or loss of data is possible. It is your responsibility to take precautions to prevent this from happening, and that means being smart with where your data is stored and how accessible it can be by an outside party. Here are a few good practices.

- Encrypt files – Documents on software programs (ex. MS Work) can be encrypted and password protected. Encryption is an excellent way of preventing access to data by unauthorized people. Be sure to keep the password stored separated from the device with the encrypted documents.
- Encrypt your laptop – FileVault is a tool available for Macs. BitLocker is available for Windows.
- Encrypt external hard drives – Always encrypt external hard drives that store research data.
- Store on the College B: drive – This is only accessible with a college ID and password. If you are off-campus, you must connect through a VPN (see section 1 above).
- Store on One Drive – One Drive is encrypted and information stored here is on Canadian servers.
- Don’t use USB Keys – USB keys are small and easy to lose. They are also generally not very stable in the long term. USB keys should be used as a last option, and they should always be encrypted.

- Password protect computer, laptop, and phone – Always password protect laptops and computers that store research data. Set a time-out as well so it automatically locks after 5 minutes.
- Watch your device – Laptops, USB keys, mobile devices, etc. should be stored safely – lock them in a drawer when not using them.

3. Transfer data properly

Sharing and sending data with other members of the research team can introduce risks. The general rule of thumb is not to transfer via the cloud, and always use secure transfer methods for sending and receiving data. Here are some good options for sharing and transferring data:

- Use College of the Rockies email only – but do not email personally identifiable participant information. If it's necessary to transfer files to another member of the research team by email, ensure the files only contain de-identified data and are encrypted and send them from College email addresses. Make sure all team members receive files directly through their institutional email (and not through a third party, like Gmail, which you should never use for College work).
- If transferring from a mobile device to computer (i.e. audio interview data), use a cable – don't email files or transfer via the internet (Cloud), and don't use syncing to transfer files from one device to another.

4. What not to use

Not all tools are suitable for storing participant research data due to storage and security risks. Do not store any personally identifying participant data on:

- Google Docs,
- Dropbox,
- Evernote,
- Box, or other cloud-based storage services.
- One Drive, if using personal health information.

5. Some good tools and tips

There are many good tools out there to help you conduct your research. These are just a few of them:

- Voice memo for iPhone – This is good for recording interviews. It is easy to prevent auto-syncing to icloud and you can keep the data on your phone.
- Dragon Dictation – this is a good tool for transcribing interviews.
- Skype for Business – the basic version is available for Windows 10 (and older versions) as a free download.

6. Make a plan that works

Be practical – Develop a plan that takes into account the security and safety of participant data, but also one that is practical and makes sense for the research team. Start by thinking what you'd like to do, and make sure each step in the process follows best electronic data security practices. There is more than one way to keep participant data secure!

Appendix 2: [Portage Data Management Plan Exemplar: Ecohydrology Research Group](#)

1. Abstract

This exemplar DMP was created by the Ecohydrology Research Group at the University of Waterloo, with the purpose of providing a standardized DMP model for its many multi- and cross-disciplinary research projects. Here, fundamental processes and procedures core to the ERG have been incorporated as a starting point for its faculty, staff, and HQPs (Highly Qualified Personnel) when developing their research projects. This exemplar demonstrates the utility of a model DMP being used by a research group or collective to maintain best practices in data management.

2. Principal Investigators

Philippe Van Cappellen, Fereidoun Reza Nezhad

3. Administrative Details

- i. Project Name:
Ecohydrology Research Group Data Management Plan
- ii. Principal Investigators/Researchers:
Philippe Van Cappellen (Orcid ID: <https://orcid.org/0000-0001-5476-0820>)
Fereidoun Reza Nezhad (Orcid ID: <https://orcid.org/0000-0002-9608-8005>)

4. Description

The Ecohydrology Research Group (ERG) at the University of Waterloo carries out fundamental research in support of the wise use of water resources, that is, one that balances society's water needs with those of natural ecosystems. ERG's research activities cover a vast range of spatial and temporal scales, from molecular-level studies on the processes determining the chemical forms and bioavailability of nutrients and pollutants to global scale assessments of anthropogenic perturbations of hydrological and biogeochemical cycles.

Faculty, staff and all Highly Qualified Personnel (HQPs – students, postdoc, research scientist/associates) in ERG are committed to data management practices to ensure that ERG data, software, code, inputs and outputs (“the data”), are safely stored, preserved and easily accessible for future re-use. ERG is aligning its Data Management Plan (DMP) with the common approach being pursued at the University of Waterloo.

In consultation with supervisors and laboratory technicians, all HQPs are expected to adopt this DMP in the conduct of their research project. They may also be asked to modify certain parts of their proposed research plans using the guidelines provided by Portage Network or to meet funding agency or contractual requirements. If, for a given HQP, significant deviations from the ERG DMP are expected, the HQP will incorporate an individual DMP in their research proposal, which will be revisited, and if needed, updated as their research project progresses.

5. Institution

Water Institute, University of Waterloo

6. Data Collection

- i. What types of data will you collect, create, link to, acquire, and/or record?

The Ecohydrology Research Group (ERG) generates data from existing documents, fieldwork, laboratory experiments, secondary data sources (including remote sensing), and numerical modelling. Data types include laboratory, modelling, hydrological, chemical, and biological data (hourly, daily, weekly, seasonal, annual) related to ERG's research goals. Additional data include field site observations/images, sampling location through Global Positioning System (GPS) coordinates/maps, climate data and modelling input files, parameters, and output files. Generated data are used in model development, calibration, and validation. Model tasks are done concurrently with the generation of the data from lab experiments and field observation within ERG. This approach allows us to discard unacceptable data and, if needed, repeat/redesign experiments.

- ii. 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?

The Ecohydrology Research Group (ERG) generates several types of primary data using various proprietary file formats from the instruments or numerical models. Highly Qualified Personnel (HQP) are expected to convert these files to non-proprietary formats, such as Comma-Separated Values (.csv) or Text (.txt) whenever possible to help facilitate future re-use.

- iii. 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?

All dates should follow [ISO 8601 date format](#) (YYYY-MM-DD) and study site names will follow the [Canadian Geographical Names DataBase](#) where applicable.

Directory and file naming conventions are established at the beginning of a project and may incorporate discipline-specific conventions for the variables collected. Version control conventions will also be established. File names include the project name (in shortened form) and a brief description of the file's content.

- For example: PX_WP1_WaterQuality_Thames_River

Document versions should be sequentially named (with file names ending in v1, v2 etc.).

- An example is the following: WSP_WaterQuality_Thames_River_raw_200617_v1

- iv. What data quality assurance and quality control methods will be used?

All incoming Ecohydrology Research Group (ERG) researchers are trained in standardized Quality Assurance/Quality Control (QA/QC) protocols by experienced technicians and staff. They are expected to implement QA/QC during sampling and sample analyses. For chemical and biological analyses, the quality control will include replicates, blanks, certified reference materials and spikes. Whenever possible, a specific lab task is to be conducted by the same student, technician or researcher for quality assurance. Final data workbooks will be stored in a project repository file server and will contain data sheets corresponding to samples and standards, Certified Reference Materials (CRMs), preparation, raw instrument output, manipulated (i.e., digestion/dilution corrected) instrument output, results, and reported limits of detection and quantification.

7. Documentation and Metadata

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

In order for the data to be read and interpreted correctly, contextual documentation is provided, including laboratory Standard Operating Procedures (SOPs), instrument operating and software manuals. Data documentation is aligned with national and international best practices.

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

All Ecohydrology Research Group (ERG) researchers will collect metadata and develop appropriate documentation at project inception with regular updates throughout the life of their research projects in accordance with practices developed by the ERG data management team. For example, when collecting laboratory and field data, all chemical nomenclature used will follow the [International Union of Pure and Applied Chemistry](#) guidelines. Researchers are asked to keep meticulous records of sampling and experimental details such as sample preparation, sampling procedures, time, etc.

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

Metadata associated with all projects is collected to aid future discovery, access and re-use. Data are described according to the [ISO 19115](#) metadata schema, with additional metadata collected per the specific needs of the Highly Qualified Personnel's (HQP) research project.

8. Storage and Backup

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

Each Highly Qualified Personnel (HQP) is expected to generate an average 500 gigabytes (GB) of data during the data management lifecycle, and all data generated for their individual project are to be stored until the thesis/article is published and a Digital Object Identifier (DOI) is assigned to the dataset to support discovery and reproducible research.

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

All Highly Qualified Personnel (HQP) are expected to follow the 3-2-1 backup and storage best practices. This means that Ecohydrology Research Group (ERG) members will create three copies of all data files, to be stored on two different types of media, with one copy kept in an off-site location.

Digital data is stored in spreadsheets, databases, word processing files and lab notebooks. Any data recorded in field or lab books are digitally transcribed, and the hard copies will be archived for a minimum of five (5) years after the life of the project.

ERG HQP decide how to best back up their data, as long as they follow the 3-2-1 rule above. However, all ERG HQP are encouraged to provide copies of their files for back up on the ERG server on a node of Science Computing at the University of Waterloo. This server is maintained and backed up regularly by Science Computing staff. These backups follow a 52-week retention schedule that is backed up weekly.

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

Active research data are stored by each Highly Qualified Personnel (HQP) responsible for its collection until it is ready for publication/dissemination. If it is necessary to share data among researchers and collaborators, the Ecohydrology Research Group (ERG) management team, with approval from the Principal Investigator/Supervisor, mediates data access and usage during the active phase of the project work. HQPs use [Microsoft Teams](#) to access and share data with collaborators. Microsoft Teams is a collaboration and communication application within the Microsoft Office Suite. Large data files may be shared via a portable drive.

9. Preservation

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

Data are to be made available within one month of publication in a trusted repository, such as the [Federated Research Data repository](#) (FRDR), that provides persistent identifiers such as Digital Object Identifiers (DOIs) or handles. Adhering to [UW Policy 73](#), all Highly Qualified Personnel (HQP) within the Ecohydrology Research Group (ERG) are also expected to provide a digital copy of all research data to their respective supervisors and collaborators at the end of their project/thesis/publication. Data will be published on the University of Waterloo [Water Institute Storage Group](#) on the FRDR, and if applicable, the Gordon Foundation [DataStream](#).

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

To facilitate interoperability, data will be saved in non-proprietary software formats which are accessible to others. Examples of these include Comma-Separated Values (.csv) for spreadsheets, Tagged Image File Format (.tiff) files for images, Georeferencing Tagged Image File Format (.geotiff) for spatial data, and Text (.txt) for text.

10. Sharing and Reuse

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

Final data used in publications and theses will be shared to demonstrate reproducibility of Ecohydrology Research Group (ERG) research.

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

Processed research data resulting in publications will be made openly available via a [Creative Commons Attribution 1.0 Generic License](#) (CC BY 1.0 License). If a more restrictive license is warranted, researchers will consult with the Ecohydrology Research Group (ERG) management team (or [University of Waterloo Library Research Data Services](#)).

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

Data deposited into the Federated Research Data Repository (FRDR) and the DataStream repository have unique Digital Object Identifiers (DOIs), searchable keywords and other searchable metadata. These metadata with data links will be made available via the Ecohydrology Research Group (ERG) website, social media and [Waterloo Metadata Index](#). ERG Highly Qualified Personnel (HQP) will also cite the data in their peer review articles. Email announcements will be sent to appropriate relevant list-servers upon release.

11. Responsibilities and Resources

- i. 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.

Each Highly Qualified Personnel (HQP) in the Ecohydrology Research Group (ERG) is responsible for the day-to-day management of their individual project data. It is expected that their overall data management aligns with best data management practices and ERG's DMP. This includes but is not limited to standard Quality Assurance/Quality Control (QA/QC) procedures, compiling, sharing, and backing up of data and metadata regularly onto secure servers. If two or more team members are working jointly on a specific research project, they will determine at the onset of their work which member is responsible for implementing the data management practices for their project. The Principal Investigators are responsible for ensuring team members follow the ERG data management plan.

- ii. 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?

Data management practices will be reviewed each semester by the Ecohydrology Research Group (ERG) management team with a minimum of two people. All researchers will be reminded of this plan through various communication channels such as email communication, ERG weekly meetings, and a [Microsoft Teams](#) communication channel.

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

Given the size of the Ecohydrology Research Group (ERG), hiring a full time data manager would be necessary to support the management of ERG research data; hopefully this type of Research Data Management (RDM) support can be built into future grant applications. However, in the interim, Highly Qualified Personnel (HQP) in the ERG are expected to leverage online resources (such as [Mantra](#) and [Portage Network Resources](#)) or seek guidance from the data manager of the Global Water Future Program and the [University of Waterloo Library Research Data Services](#) to ensure data management best practices are followed.

12. Ethics and Legal Compliance

- i. 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?

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 [University of Waterloo Office of Research Ethics](#). Only appropriate and approved metadata will be made available. Highly Qualified Personnel (HQP) with sensitive data are encouraged to ensure that files are encrypted.

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

Any legal, ethical, and intellectual properties issues will be guided by relevant policy(ies) of the University of Waterloo and the Laws of Canada.

Appendix 3: Portage Data Management Plan Guidance: Qualitative Mixed Methods

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.

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)

1. Data Collection

- i. Describe the type(s) of data that you will collect

Include 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.

- ii. 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.

Examples 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.

iii. 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.

Surveys:

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).

Transcription:

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.

iv. 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.

v. 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:

a. Surveys:

- CSV;
- HTML;
- Unicode Transformation Formats

b. Qualitative Interviews

Audio Files:

- MP3;
- FLAC;
- Ogg

Video Files:

- MP4;
- .mkv

c. 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.

2. Documentation and Metadata

- i. 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>.

- ii. 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](#).

- iii. 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.

- iv. 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.

3. Storage, Backup, and Access

- i. 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.

- ii. 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](#).

- iii. 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](#).

4. Preservation

- i. 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.

- ii. 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](https://www.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](#).

5. Sharing and Reuse

- i. 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.

- ii. 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.

iii. 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](#).

6. Responsibilities and Resources

- i. 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.

- ii. 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.

- iii. 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.

7. Ethical and Legal Compliance

- i. 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.

- ii. 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.).