

Data Management Plans

INTRODUCTION TO DMPs

A Data Management Plan (DMP) is foundational to successfully manage your research data. A DMP is a brief formal document that details the strategies and tools you will use to manage your data throughout the course of your project and long after its completion and/or publication.

Created alongside your research plan, a DMP is a “living” document that should be reviewed, modified, or updated throughout your project to accurately reflect your data management practices.

Creating a DMP helps you in many ways:

- Fulfills grant or publisher requirements
- Facilitates the organization of your data to make it easier for your research team to find, use, and document the project data and files
- Identifies your needs for storing, documenting, and managing your research data early
- Budget costs of data storage and management to include in your grant proposal
- Makes your research outputs more reliable, accurate, and reproducible
- Makes your data FAIR (findable, accessible, interoperable, and reusable) for maximum impact

Sign up with the [DMP Assistant](#) to write your plan

The Digital Research Alliance of Canada has created [video tutorials for the DMP Assistant](#) for more guidance on how to access and use the tool.

HOW TO WRITE A DMP

These sections follow the general Portage Template, available through the DMP Assistant, and highlight examples from the [DMP Exemplars](#) from the [Digital Research Alliance of Canada](#) (formerly Portage Network). The templates available through the DMP Assistant include additional guidance for each question/section of the DMP.

Project Details

Title: DMP Guide with Examples

Creator: Sarah Giesbrecht

Affiliation: The University of Winnipeg

Template: Portage Template

Project abstract:

This Data Management Plan uses responses from the various DMP Exemplars created in partnership with Portage Network. This is intended to be a multi-disciplinary guide to writing DMPs using the DMP Assistant.

Last modified: 26-07-2022

Data Collection

What data will you collect?

Data can be quantitative survey data, digital audio files, written transcripts, photographs, textual data for qualitative analysis, geospatial data, and many other types. Just be specific when describing your data.

Examples:

We will be collecting bibliographic metadata about titles of written works, persons and firms such as title of work, name of author, publisher, date published, etc. Currently we are collecting data from about 50 different sources, some print, most digital. Examples include the British Library online catalogue and specialized print bibliographies like the English Novel 1770-1830.

Source: Levy, M. (2020). *Data Management Plan Exemplar #1: Digital Humanities - Women's Print History Project (1750-1830)*. <https://doi.org/10.5281/zenodo.4064168>

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.

Source: Persaud, B., Van Cappellen, P., Reza Nezhad, F., Neilson, M., & Szigeti, K. (2020). *Data Management Plan for Ecohydrology Research Group (Exemplar)*. <https://doi.org/10.5281/zenodo.4062478>

What file formats will your data be collected, analyzed, and stored in?

These will often depend on the software you plan to use. Try to use open source file formats or industry-standard formats (e.g. those widely used by a given community) to prolong the lifespan and usability of your data.

- Statistical data files e.g. SPSS, SAS, Stata, R, NVivo
- Spreadsheets e.g. Excel or Google Sheets
- Text data e.g. .txt text files, .csv comma separated values
- Documents e.g. MS Word, Google Docs
- Image files e.g. tiff, jpeg

Example:

Soundscapes will be recorded in WAV (Waveform Audio File Format) and then migrated to BWF (Broadcast Wave Format), which allows for richer metadata capture. This file format is recognized as a preservation format by the International Association of Sound and Audiovisual Archives, among others, and is

recommended in a document produced by the Canadian Heritage Information Network in collaboration with members of the Digitization and Digital Preservation Discussion Group and presented to the National Heritage Digitization Strategy (NHDS) Steering Committee. The same document recommends a sampling rate of 96 kHz and 24-bit coding. Since BWF is backward compatible with WAV, only BWF files will be retained.

Source: Paquette-Bigras, È. (2020). *Data Management Plan: Soundscape Study (Exemplar)*.

<https://doi.org/10.5281/zenodo.4056785>

How will you name and organize your files and keep track of the different versions as your project progresses?

It is important to agree upon a naming convention early in your project to reduce confusion for you and your research team and help prevent data loss as your project progresses. Generally, use a short, unique, and descriptive name or identifier, use the year-month-day standard for dates (YYYY-MM-DD), and include the version number in your file names.

For information on file naming and organization, visit the [Documentation and Metadata Best Practices](#) guide.

Examples:

Each file will be named with a short description/acronym to reflect its content, followed by the date of creation. To record different versions, we will add a version number in the file name. For example, file name GSC_20200608_v01.xls represents the data about Google Scholar Citations acquired on June 8, 2020, the 1st version.

We will create a document to detail file naming conventions and provide a list of explanations of the short descriptions/acronyms used in file names.

Source: Kumaran, M., & Zhang, L. (2020). *Data Management Plan for Usage of Academic Profile Websites (Exemplar)*.

<https://doi.org/10.5281/zenodo.4062489>

Documentation and Metadata

Metadata and documentation add labels and information to your data to make it usable and understandable. The more descriptive metadata you include, the easier it will be for you and your team to keep track of your data as well as make it possible for others to reuse your data if you share it.

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

Your documentation should include information on when and how the data were created/collected and how data files are stored, structured, and modified. You should also include a license and any restrictions on use.

Your data documentation may also include a “readme file”, which lists the contents of files or folders and explains what is in each file. This document describes variables and explains how variables, columns or other data elements are named, and explains any codes and what they correspond to.

For more information on ReadMe files, visit the [Documentation and Metadata Best Practices](#) guide.

Examples:

We will provide a brief description of the project, a detailed methodology on how the data is collected, date of data collection, analysis performed, and details of who performed each task to accompany the data. A data dictionary will be created to define different data fields, data type, and explanation of data coding.

Source: Kumaran, M., & Zhang, L. (2020). *Data Management Plan for Usage of Academic Profile Websites (Exemplar)*. <https://doi.org/10.5281/zenodo.4062489>

A "README.txt" file will accompany the dataset. It will contain general information on the dataset, information on how the dataset is shared, an overview of the files it contains, information on the research methodology, and information on the data, such as the number of variables or the code for missing data. Each recording file will also be accompanied by an XML (Extensible Markup Language) file containing metadata.

Source: Paquette-Bigras, È. (2020). *Data Management Plan: Soundscape Study (Exemplar)*. <https://doi.org/10.5281/zenodo.4056785>

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

In addition to good research data management practices, information about your data collection and methodology will be needed in your final publication or for grant applications, so take the time to plan and capture documentation early about the overall study, your data collection processes, and information needed to understand individual files, to save time when you will need this information later.

Example:

- 1. A working document will be created and revised as new content is added to the database.*
- 2. The field structure is stored within Inmagic and may be printed out as a text file.*
- 3. Data entry instructions are included on the web-based data entry form within Inmagic and may be printed out as a text file.*
- 4. These lists are stored within Inmagic and may be printed out as text files.*

Source: Gray, V., & Cooper, A. (2020). *Data Management Plan Exemplar #2: Digital Humanities and Secondary Data*. <https://doi.org/10.5281/zenodo.4019309>

If you are using a metadata standard and/or tools to document and describe your data, describe the metadata standard/tools you will use.

Some disciplines have required Metadata standards particularly for some scientific data but a specific standard may not be necessary for your project. You may want to create your own metadata standard to ensure metadata stays consistent within your team throughout your project.

Example:

We have developed our own metadata standard appropriate to the nature of the project. The description of our metadata standards, which also provides contextual information regarding how the tables are linked, is provided as an Appendix to this document.

Source: Levy, M. (2020). *Data Management Plan Exemplar #1: Digital Humanities*.
<https://doi.org/10.5281/zenodo.4064168> See Appendix pages 9-13

Data Storage and Backup

For more information on Secure Data Storage and Backups, visit this [Secure Storage](#) guide.

How much space do you anticipate needing to store your data and for how long?

Do you anticipate needing more storage space than previous research projects?

Example:

Storage space is anticipated to be approximately 100GB. The data will be stored for 5 years locally, with a permanent copy held in the Scholars Portal Cape Breton University Dataverse.

Source: Leviten-Reid, C. (2020). *Data Management Plan for People, Places, Policies and Prospects: Affordable Rental Housing for Those in Greatest Need (Exemplar)*. <https://doi.org/10.5281/zenodo.4062466>

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

If you are unsure how to keep your data secure and backed-up, please consult with Technology Solutions Centre and/or the Research Data Management Librarian especially if you will need to store your data on University servers or the UWinnipeg Research Data Repository at any point during or after the course of your project. If applicable, also describe where physical samples and hardcopies will be securely stored. For more information on Secure Data Storage and Backups, visit this [Secure Storage](#) guide.

Example:

Data will be stored on the principal researcher's hard-drive and shared institutional Google drive with regularly scheduled uploads to Dataverse, the University of Alberta's data repository. Data will also be uploaded using Web FTP or file transfer protocol to a Web publishing platform, such as Omeka or WordPress, maintained by UA Arts Resource Centre (ARC) where additional metadata will be incorporated. Images and geographic data will also be uploaded to Google maps, which in turn will provide a .kml/.kmz file for the preservation of the maps created for this project. Bibliographic data will be stored in RefWorks and exported in RIS/Bibtex format for preservation. Data stored in Dataverse and the public website created for the project will be permanently preserved by the UA Library and Archivelt respectively.

Source: Lacroix, D., & Rao, S. (2020). *Data Management Plan for Belgians and French in the Prairies (Exemplar)*.
<https://doi.org/10.5281/zenodo.4062484>

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

You will want a storage solution that is easy to use and allows for collaboration but also ensures data security. Finding a solution that fits these three requirements becomes increasingly challenging when a research team has members outside of just one institution. Seemingly simple solutions such as relying on email or third-party

commercial services (such as Google Drive and Dropbox) are not necessarily secure and robust. Contact the Research Data Management Librarian for advice on secure and collaborative data storage solutions.

Example:

The 3-2-1 backup rule will be followed for data storage and backup. This means that team 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. Examples of different media types include those which are removable (USB), fixed (such as a hard drive on a laptop) and networked (such as cloud-based servers). Team members can best decide how to back up their data, as long as they follow the 3-2-1 rule above and that it aligns with any institutional and/or ethical requirements. However, all team members will upload their files to a cloud-based server located in Canada, to be identified by the project lead. Sensitive files are to be encrypted.

Source: Leviten-Reid, C. (2020). *Data Management Plan for People, Places, Policies and Prospects: Affordable Rental Housing for Those in Greatest Need (Exemplar)*. <https://doi.org/10.5281/zenodo.4062466>

Preservation

Before answering the following questions, consider **how you will select which data to preserve** and **how long you want to preserve your data**. Your grant or publisher may have specific time requirements. Although these questions are not included in this version of the DMP Assistant template, including this information will strengthen your DMP.

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

Contact the UWinnipeg Research Data Management Librarian to discuss various data repositories and deposit options to find which is best suited for your data.

Example:

The project does not have a foreseeable end date. An ASCII delimited (and potentially a XHTML) version of the database will be created and could be stored on Scholars Portal Dataverse, a data repository which assigns DOIs to datasets, and supports preservation, discovery, citations, and data usage metrics. However, a consultation with the University's Research Data Management Librarian will help identify other possible repository options for our research data.

Source: Gray, V., & Cooper, A. (2020). *Data Management Plan Exemplar #2: Digital Humanities and Secondary Data*. <https://doi.org/10.5281/zenodo.4019309>

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.

Example:

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.

Source: Persaud, B., Van Cappellen, P., Reza Nezhad, F., Neilson, M., & Szigeti, K. (2020). *Data Management Plan for Ecohydrology Research Group (Exemplar)*. <https://doi.org/10.5281/zenodo.4062478>

Sharing and Reuse

What data will you be sharing and in what form?

E.g. raw, processed, analyzed, final

- the raw data from your initial observations or survey
- data that has been processed or cleaned to remove errors and made the data ready for analysis, or processed to de-identify human participants
- data that has been processed for preservation purposes and reformatted in a preferred preservation format
- summary data that has already been manipulated and analyzed

Example:

The raw data (i.e., bibliographic and holdings information) will be shared.

Source: Gray, V., & Cooper, A. (2020). *Data Management Plan Exemplar #2: Digital Humanities and Secondary Data*. <https://doi.org/10.5281/zenodo.4019309>

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

Your publisher or grant may have specific end-user license requirements. If you are using third-party data, the original data compilers may also have licensing requirements. We recommend using one of the standard licenses available to researchers, such as the [Creative Commons licenses](#) and the [Open Data Commons licenses](#).

Example:

We will have the Creative Commons Attribution CC BY license for the data, which allows others to distribute, reuse, adapt, and build upon the data as long as the original data creators are credited.

Source: Kumaran, M., & Zhang, L. (2020). *Data Management Plan for Usage of Academic Profile Websites (Exemplar)*. <https://doi.org/10.5281/zenodo.4062489>

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

Publishers will often ask for a DOI or URL so they can link to your data in your article.

Example:

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.

Source: Persaud, B., Van Cappellen, P., Reza Nezhad, F., Neilson, M., & Szigeti, K. (2020). *Data Management Plan for Ecohydrology Research Group (Exemplar)*. <https://doi.org/10.5281/zenodo.4062478>

Responsibilities and Resources

Who will be responsible for managing this project's data during and after the project? What are the major data management tasks for which they will be responsible?

It is also good practice to outline roles on responsibilities of each member of the research team and include the responsibilities of the Research Data Management Librarian if you plan to use the UWinnipeg Research Data Repository or other Research Data Management services or resources.

Example:

The principal investigator will be responsible for data management during and after the project. The Principal Investigator will ensure that those assisting her – when she is working with assistance – follow appropriate research data management practices. Tasks performed by the main investigator and those assisting her include data collection; data transfer, storage, and backup; data processing; project and data documentation; and data analysis.

Source: Paquette-Bigras, È. (2020). *Data Management Plan: Soundscape Study (Exemplar)*. <https://doi.org/10.5281/zenodo.4056785>

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?

Example:

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.

Source: Persaud, B., Van Cappellen, P., Reza Nezhad, F., Neilson, M., & Szigeti, K. (2020). *Data Management Plan for Ecohydrology Research Group (Exemplar)*. <https://doi.org/10.5281/zenodo.4062478>

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

Depending on the size of your project and the amount of data to be managed, there may be additional storage, software, and personnel costs at various stages of your project. Consider the financial needs of your data early on so you can include these costs in your grant proposal.

Examples:

Storage during research is provided free of charge by Compute Canada and by the principal investigator's institution on One Drive. Publication of metadata in the institutional Dataverse repository is also complimentary. The principal investigator also has an external hard drive.

Source: Paquette-Bigras, È. (2020). *Data Management Plan: Soundscape Study (Exemplar)*. <https://doi.org/10.5281/zenodo.4056785>

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.

Source: Persaud, B., Van Cappellen, P., Reza Nezhad, F., Neilson, M., & Szigeti, K. (2020). *Data Management Plan for Ecohydrology Research Group (Exemplar)*. <https://doi.org/10.5281/zenodo.4062478>

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?

This is an important question to consider even for **non-sensitive data** as the more people with access raises the risk of data loss. See the [Research Data Storage and Security Guide](#) for more information. If you are working with Indigenous community data, ensure you follow their protocols so they will have [ownership, access, control, and possession](#) of their data.

Examples:

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.

Source: Persaud, B., Van Cappellen, P., Reza Nezhad, F., Neilson, M., & Szigeti, K. (2020). *Data Management Plan for Ecohydrology Research Group (Exemplar)*. <https://doi.org/10.5281/zenodo.4062478>

Some of our data may be sensitive data and will be placed on secure cloud servers, where it will be encrypted. Only analyzed, de-identified data will be made available once the project is complete.

Source: Leviten-Reid, C. (2020). *Data Management Plan for People, Places, Policies and Prospects: Affordable Rental Housing for Those in Greatest Need (Exemplar)*. <https://doi.org/10.5281/zenodo.4062466>

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

If you wish to share de-identified or anonymized data after the completion of your project, include information on secondary uses of the data in your consent forms. For example language and more

information, see this guide to [Human Participant Data](#) and review the [UWinnipeg Research Ethics Guidance Documents](#). If you are using Indigenous data, follow the data sharing protocols of the community represented in the data.

Examples:

Although there is no sensitive data involved in this project, we will have names, rank, and institution information of the study objects. Releasing the individual names publicly may cause unnecessary concerns. Therefore, we will replace the names by study object number, such as Librarian Researcher 1, 2, 3, etc., when depositing data in a repository when the project is completed.

Source: Kumaran, M., & Zhang, L. (2020). *Data Management Plan for Usage of Academic Profile Websites (Exemplar)*.

<https://doi.org/10.5281/zenodo.4062489>

No sensitive data will be shared. De-identified and non-sensitive data will be made available on the Scholars Portal Cape Breton University Dataverse. Any sensitive data will be stored on secure servers for 5 years.

Source: Leviten-Reid, C. (2020). *Data Management Plan for People, Places, Policies and Prospects: Affordable Rental Housing for Those in Greatest Need (Exemplar)*.

<https://doi.org/10.5281/zenodo.4062466>

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

For guidance, visit the [UWinnipeg Research Ethics website](#), the [Privacy Office Resources](#), and the [Copyright Office website](#).

Example:

Participants have signed a consent form authorizing the public domain release and reuse of their soundscapes, as well as a copyright assignment and moral rights waiver form.

Source: Paquette-Bigras, È. (2020). *Data Management Plan: Soundscape Study*

(Exemplar). <https://doi.org/10.5281/zenodo.4056785>

SUBJECT SPECIFIC DMP TEMPLATES

If the template above is not detailed enough for your field of research, the following are subject specific DMP Templates for a variety of different disciplines, which are also available using the DMP Assistant, including:

- [Advanced Research Computing](#)
- [Arts-Based Research](#)
- [CRDCN Template for Accessing Data from Research Data Centres](#)
- [CRDCN Template for Research Data Centres and External Analysis](#)
- [History and the Humanities](#)
- [Interdisciplinary Health Software/Technology Development](#)
- [Mixed Methods \(Surveys & Qualitative Research\)](#)
- [Neuroimaging in the Neurosciences](#)
- [Open Science Workflows](#)
- [Qualitative Health Sciences Research](#)

- [Studying Molecular Interactions](#)
- [Systematic Reviews](#)
- [Water Quality Research](#)