University of Windsor: University of Windsor

Data Collection

What data do you intend to collect?

Guidance:

Describe the data you intend to collect. Examples include: images, audio, video, textual for qualitative analysis, survey data for quantitative analysis, modeling data, spatial data, instrumentation data.

What file formats will your data be collected and saved in?

Guidance:

Proprietary file formats which require specialized software or hardware to use are not recommended, but may be necessary for certain data collection or analysis methods. Using open file formats or industry-standard formats (e.g. those which are widely used by a given community) is preferred whenever possible.

Software-specific (proprietary) formats include: SPSS .sav file, Matlab .mat file, Stata .dta file; image filetypes (tiff, jpeg); spatial filetypes (.shp); textual (.txt, .xml). Data may also be stored in a database (e.g. RedCap) or in an instrument-specific format.

Read more about file formats on the UK Data Archive website.

How much data storage space do you anticipate needing for the data you will collect?

- Less than 1 TB
- 1 - 10 TB
- 10 - 100 TB
- More than 100 TB

Guidance:

The type of data you collect will impact your storage requirements. 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.

If your data will be changing frequently, please estimate how much extra storage will be required to store multiple versions. If you are collecting data over a long period of time (e.g. several months or years), please describe your plan for adding additional storage to accommodate the growth of your data.

Will you be re-using any data? If so, please explain how you will obtain that data and integrate it into your research project.

Guidance:

Include data you or your research team has collected previously, public data, and data from other researchers.

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

Guidance:

Ensuring that you and your research team are using the 'master' version of your data is extremely important. Establishing conventions for naming and versioning your files is essential, especially when data may be available in different file formats or stored on different computers.

More information on file naming and version control is available from the University of British Columbia and UK Data Archive.

University of Windsor's Borealis can help with version control of your files.

Documentation and Metadata

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

Guidance:

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 require to reproduce derived data.

More information about data documentation is available at the University of British Columbia.

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

Guidance:

Metadata is a description of the contents and context of data files. Using a metadata standard (a set of required fields to fill out) helps 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 general and discipline-specific metadata standards and tools for research data. For assistance with choosing a metadata standard, please contact libdata@uwindsor.ca.

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

Guidance:

It is best to begin the documentation process at the start of a research project. Details or procedures may be forgotten if documentation is created at the end of a project.

Storage and Backup

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

Guidance:

Storing data on flash drives, laptops, computers, or external hard drives without a regular backup procedure is not recommended, because they are subject to hardware failure, viruses, file corruption, theft, loss, and human error. Likewise, multiple copies of data stored in the same physical location does little to mitigate risk.

Using software which automatically backs up your data and keeps it synchronized with a master copy is recommended. Copies of data should ideally be stored in separate physical locations to mitigate physical risks, e.g. fire.

Local active storage options include UWindsor's OneDrive as well as secure departmental networks which are regularly backed up.

University of Windsor also offers Borealis for storing and sharing small datasets. For assistance with Dataverse, contact libdata@uwindsor.ca.

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

Guidance:

The security measures you choose to implement will depend on the sensitivity of the data being stored. Data can be protected by limiting physical access to hard drives and file servers on which it is stored, by storing data on computers that do not have external network access (i.e. access to the Internet), through password protection, and by encrypting data files.

Sharing sensitive data via email and cloud storage services such as Dropbox is not recommended, since the security of these services cannot be trusted.

Preservation

Which data will be selected for preservation?

Guidance:

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

Will you deposit your data for long-term preservation at the end of your research project? If so, where?

- Data will be retained locally (OneDrive, network drive)
- Data will be deposited in UWindsor's Dataverse
- Data will be deposited in FRDR (Federated Research Data Repository)
- Other - please specify

Guidance:

The University of Windsor's Dataverse (hosted by ScholarsPortal at the University of Toronto) is available to UWindsor researchers and can serve preservation needs where file sizes are less than 3 GB. Dataverse allows researchers to restrict access to datasets in compliance with research ethics guidelines, but is not suitable for highly sensitive data.

Another option is the Federated Research Data Repository (FRDR) which allows for deposit of files up to 5 TB. FRDR allows researchers to set an embargo period for their data.

For larger datasets or help identifying other data repositories, please contact libdata@uwindsor.ca.

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

Guidance:

It is likely that you will have multiple versions of your data (raw, processed, analyzed, final), and your data may be stored in different file formats depending on how they were collected and analyzed.

Digital information may be lost while converting files from one format to another (e.g. converting an image from a non-compressed 'lossless' TIFF file to a compressed 'lossy' JPG file), so changes to file formats should be clearly documented.

Identify what steps are needed after your project is completed in order to ensure the data you are choosing to preserve or share is anonymous, error-free, and converted to recommended formats with a minimal risk of data loss.

Data Sharing and Reuse

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

How will you be sharing your data? (e.g. institutional repository, a specialized data archive, informal/on-request sharing).

- University of Windsor's Dataverse repository
- Federated Research Data Repository (FRDR)
- Journal-specific repository (specify below)
- Figshare
- Other public repository (specify below)
- By request only
- Other (specify below)
- Unable to share

Guidance:

Where you share your data depends on the complexity and size of your data. Smaller, low-risk datasets can be shared via institutional or disciplinary repositories and project websites. Larger datasets or datasets with more stringent security requirements cannot be easily transferred over the Internet.

Researchers at the University of Windsor can deposit smaller datasets (less than 500MB) in Borealis (for complex data formats such as that produced by SPSS, SAS, Stata, R or basic data formats such as Excel/CSV. For larger datasets, please contact libdata@uwindsor.ca.

Please describe whether there will be any restrictions placed on your data when they are made available and who may access them. If data are not openly available, describe the process for gaining access.

Guidance:

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
- gaining participant consent for data sharing
- gaining permissions to share adapted or modified data
- and agreeing to a limited embargo period.

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

Guidance:

Assigning an end-user license allows you to define the terms of use for your data. We recommend a Creative Commons license.

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?

Guidance:

Consider where, how, and to whom sensitive data with acknowledged long-term value should be made available, and how long it should be archived. These decisions should align with Research Ethics Board requirements. The methods used to share data will be dependent on a number of factors such as the type, size, complexity and degree of sensitivity of data.

Outline problems anticipated in sharing data, along with causes and possible measures to mitigate these. Problems may include confidentiality, lack of consent agreements, or concerns about Intellectual Property Rights, among others. In some instances, an embargo period may be justified; these may be defined by a funding agency's policy on research data.

Restrictions can be imposed by limiting physical access to storage devices, by placing data on computers that do not have external network access (i.e. access to the Internet), through password protection, and by encrypting files. Sensitive data should never be shared via email or cloud storage services such as Dropbox.

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

Guidance:

Obtaining the appropriate consent from research participants is necessary 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. For example, it may stipulate that the data will only be shared for non-profit research purposes or that the data will not be linked with personally identified data from other sources.

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

Guidance:

Compliance with privacy legislation and laws that may impose content restrictions in the data should be discussed with your institution's privacy officer or research services office. Research Ethics Boards are central to the research process.

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.).

Responsibilities and Resources

Who will be responsible for data management during the project? (i.e. during collection, processing, analysis, documentation). Identify staff and organizational roles and their responsibilities for carrying out the data management plan (DMP), including time allocations and training requirements.

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

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

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