UNIVERSITY OF TORONTO LIBRARIES

Data Management Plan (DMP) - Question Guide

What is a DMP? A DMP (Data Management Plan) is a formal plan that outlines how a researcher intends to manage research data during and after a research project. **Why do I need a DMP?** Many funders strongly encourage or require researchers to complete a DMP to help ensure the appropriate management of research data for reproducibility and reuse by others. Typical DMP questions are presented below with assisting guidance. Note that some funders may require a specific DMP template be used, always check first. For more information to help you with creating a DMP please refer to the UTL RDM website or contact rdm@utoronto.ca for support.

DMP Section	DMP Questions	Guidance
Section 1: Administrative Information	1.1 Who is responsible for the data and who can access it? *Who is the main researcher(s) Principal Investigator(s), or Co-Investigator(s) for the study?* *Is the study being conducted by only you, or are there members on a team? If others, who is on the team (grad students, postdocs, researchers from other institutions, etc.)* *Who has access and permission to use and edit the data?* *Have you considered who will be the ongoing main contact for the data?* *Who is responsible for answering questions about the data?* *Who is responsible for granting access to the data?*	Consider who will create the data, including individuals and/or teams involved, whether they are physically located where the data are being stored or elsewhere, and, who will need to access or edit this data either immediately, or, at different stages of the research process including during and after the project has been completed. Preparing for data management: ● Assign roles for data management responsibilities; ● Appoint a main contact for the data should there be inquiries or requests for the data; ● Designate a backup contact if this individual leaves the project or the institution; For questions about getting started with data management planning contact rdm@utoronto.ca.
Section 2:	2.1 What types of data are being used,	Data may be defined as observational, experimental, simulation, models, etc. This

Data Types and Data Collection	collected, or created? *What is the nature of the data in terms of the end-product research (e.g. interviews, experimental, observational)?* *What type of data and formats do you have?* *Are you using data from another source? What sources?* *Are there any privacy or sensitive considerations involved?*	can include data from other sources (secondary data). Examples of data types include: text, spreadsheets, images, 3D models, software, audio files, video files, surveys, patient records, specimen samples, sensor readings, results from theoretical calculations, etc.[1] Sensitive Data If applicable, data collected have additional considerations such as participant consent, ethical, and privacy related regulations that may impact data management decisions. Consider mechanisms to manage sensitive data or confidential data (secure storage, and encryption or anonymization, if data will be kept and/or accessed by others at any point). For more information about managing different types of data, including sensitive data refer to the UTL RDM website. Whenever planning and conducting research, please consult UofT's Research Policies and Guidelines. The Research Ethics Board (REB) maintains useful Data Security Standards and protection guidelines for you to follow to manage sensitive and confidential data.
Section 3: Metadata & Documentation	3.1 What supporting documentation is required to understand the data? *Will someone who is unfamiliar with the project / study be able to understand the data?* *How are the data captured? Is this documented?* *Is any special software required to read the data and/or documentation?* *How are the data organized? How are the data files and folders structured? Is there a readme file or equivalent?*	Detailed descriptions accompanying data can include: code lists, log books, readme files, analysis notes, coding and syntax, data dictionaries, codebooks, thesauri, data vocabularies, classifications, ontologies, user guides, instructions, questionnaires, etc. Describe your data consistently with structured documentation so others can easily understand and reproduce or reuse your data. Data require: ● Accurate documentation describing the steps taken to process and code data; ● Meaningful labels and descriptions at various levels; ● Avoid the use of acronyms and other ambiguous language. More information is available from the UTL RDM website.

[1] See more data types: http://www.science.gc.ca/eic/site/063.nsf/eng/h_97609.html#1a

	3.2 What metadata standard(s) are used to describe the data? *Is there a discipline-specific standard that can be used?* *How will the standard metadata be generated? Will this be in a tool or specialized software?*	Metadata standards provide a set list of descriptive fields (similar to a vocabulary) for providing relevant contextual information to describe data. Most standards use open and machine-readable formats such as JSON or XML. This can improve future exchange and reuse of metadata between systems, including software used for data analysis, and, for data sharing and indexing in an online database for search and discovery purposes (e.g. Google). More information about general and disciplinary metadata standards is available from the UTL RDM website.
		Discipline-Specific Metadata There are defined disciplinary specific metadata standards that are widely used to describe research data in a particular field or domain. Discipline standards assist with domain understanding and interoperability between datasets in a particular field (e.g., the use of certain terms to describe a data collection methodology or characteristic about data from that field). General standards also exist, often with less descriptive granularity, that can be helpful for describing many kinds of data at a high-level. More information about metadata is available from the UTL RDM website.
Section 4: Storage & Security	*4.1 Who is the data manager and/or designated data users among your team?* *How will you and/or the research team and others access, modify, and contribute data throughout the project?* *Does the data have personal or identifiable confidential information?* *Are there any other data security considerations?*	If you are part of a research team where multiple people have access to edit data, it is important to know who has access and how changes to the data are being tracked. Tips for managing changes to data: • Assign a Data Manager to track data access and monitor activities (e.g. versioning); • Document any edits to the data using file versioning (some systems can do this for you or assist with it; date stamps and versioned file names are also helpful), to avoid issues with version control; • Keep a master copy of the data that only select individuals have access to. Access to reliable and safe storage during the project is required. Try to document

		the length of time this storage will be available for use as well as if there are plans to migrate the data in the future. More information about versioning and data storage is available from the UTL RDM website.
	4.2 Are you working with or sharing only the de-identified or non-confidential data? *[If working with sensitive or confidential data] Are you keeping the data in a secure server environment with access only via virtual private network or encryption operations?* *[If using encryption operations] Are you storing the encryption key(s) separately from the data? Who has access to the key(s)?*	Securing sensitive or confidential data Storage and security of data should always be considered especially if the data are confidential and/or sensitive in nature *(e.g. social insurance numbers, personal addresses, health records, etc.)*. Tips for securing sensitive or confidential data: ● Host sensitive data on infrastructure physically located at the University or on Canadian networks and servers at all times; ● Avoid working directly with sensitive or confidential data, if you must work with this data do so in accordance to and with prior approval of the UofT REB[2]. ● Investigate data anonymization[3] and encryption[4] operations; ● Always ensure the storage options chosen support the safe handling of data. More information is available from the UTL RDM website.
	4.3 *Where will the research data be stored both in the short term for the project and long term once the project is completed?* W*hat is the backup and retention schedule for the data?*	Storage and backup choices depend on the status of the data collection, data type, sensitivity of the data, etc. If the project is active, you may consider short term storage solutions that support active research (such as versioning or collaboration). Longer term options for completed projects may be required for your research data and will have policies that cover responsibility for long-term access and preservation. Always consider storage that has built in backup and synchronization options.

[2] The Research Ethics Board (REB) maintains useful Data Security Standards and protection guidelines for you to follow to manage sensitive and confidential data.
[3] Anonymization, the act and process of removing personally identifiable data or information so that at no point can it be reversed to identify an individual.
[4] Data encryption, the technical term used to describe the random codification or scrambling of data to prevent others from understanding or reading without access permission.

	Will you be using cloud storage for the long-term? Or will this be held on physical media such as a laptop/desktop or USB key? *Are the data being locally hosted or is this an external service?*	Ensure the following: ● Storage and backup of at least three copies of your data at any given time; ● Store and backup in different storage locations, for example use institutional file storage, cloud storage, or even e-mail copies to yourself frequently. ● A USB and/or personal computer or harddrive should never be the only storage for your data. More information about data storage is available from the UTL RDM website.
Section 5: Sharing & Reuse	5.1 *Are you planning to publish and share the data?* *Are there any reasons that the data cannot be shared?* *Are there any issues with sharing related to intellectual property rights, proprietary data, or legal or ethical restrictions?* *Is there a delay required for the publication of the data?* *Can anything be done to mitigate any restrictions to allow publication of all or parts of the data to be published?*	Sharing data is important for research reproducibility and reuse. Sharing allows others to use and cite your data which can increase the impact of your research. However, some data simply cannot be shared for contractual, legal, or ethical reasons. Tips for evaluating and/or preparing data for sharing: ● Detail any reasons that sharing your data may need to be restricted or delayed; ● Consider how to anonymize data if it contains sensitive or confidential information before sharing; ● Include any terms or provisions for how data can be shared within certain limits (e.g. data use applications, terms of use, access conditions and criteria for only authorized uses, etc.). More information is available from the UTL RDM website.
		Sharing Sensitive Data If you are gathering sensitive data (e.g. human subject data) that you intend to share, you will need to obtain permission (consent) from the participants in your study. The consent form for the study should detail any data sharing intentions, including data publication and data retention plans, and should also take into account protections such as access restrictions and any anonymized data outputs planned. Without consent from participants, data may not be legally shared. More information about informed consent and sample forms can be found on the UTL RDM website, or through the Office of Research and Innovation and Research

		Ethics Board (REB)[5]
	5.2 How will others find and access the data? *Which data will be shared and in what format(s)?* *Will the data be published and shared openly? If so, how? If not shared publicly, how will access be granted?* *What sort of end-user licence will you include with the data that is shared to stipulate conditions of use? (e.g. Creative Commons)*	Provide a description of the specific data that will be shared, as well as the formats they will be in. Tips for making data accessible and reusable: ● Share data in commonly used and non-proprietary formats; ● Consider packaging data, documentation, and any code together (including programming scripts, software, tools, etc.); ● Link to and cite data from any associated publications; ● Provide a brief standard description of the data and a set of keywords for your data (consider future uses of the data by others); ● Describe how others will access the data in the future (be aware of or evaluate any policies or costs associated with deposit and access); ● Ensure that the selected platform for sharing allows the data to be searchable in relevant search engines, indices or databases. The University of Toronto Dataverse is an open and free data repository option for sharing data at no cost to researchers. Other repository options include: ● Dryad, Zenodo, or Figshare; ● You can use re3data.org to search for repositories in your discipline/area of research. More information about data deposit is available on the UTL RDM website. Contact rdm@utoronto.ca with any questions about where to deposit your data.
		Licensing and Terms of Use Determine any restrictions or conditions you will place on the use or dissemination of the data shared, and how you will communicate these restrictions. *Creative Commons:*

[5] UofT REB maintains useful Data Security Standards and protection guidelines for you to follow to manage sensitive and confidential data.

		• Provides common licensing terms for reuse permissions and attribution (citation) requirements; • Can be applied to open data, for example Dataverse offers different CC licensing options (the default license is CC-Public Domain - Open reuse, no attribution required); • Specify CC-BY if you require attribution[6]. Always consider the support available in the repository to assign licensing and restrict data that require it.
Section 6: Retention & Preservation	6.1 What happens to the data once the project ends? *Will the data be kept or destroyed?* *If kept, for how long? Does this include all the data (raw or summary data) or only certain files?* *Consider whether any sensitive data should be kept, if so how will it be managed for long-term safe keeping?* *Where will the data be archived/kept? Who is responsible?* *What data format(s) will the data be stored in and what associated files and documentation will accompany the data for long-term reuse?*	Preparing data for long-term retention and archiving can be time consuming and resource intensive. It will be easier to plan for this ahead of time, consult the library about data deposit and archiving services. Planning for long-term archiving of data: • Consider if all the data are to be kept or if some or all will be destroyed once the project is finished; • Consider any standards and formats that will be used to store documentation, data, or code for future reuse; • Consider where the data will be stored long-term, and who will be responsible for it. Open standards for data and metadata (e.g. ascii text based formats, or formats that are not tied to a closed/proprietary software), tend to be much more stable over time. More information is available on the UTL RDM website. Contact rdm@utoronto.ca with any questions about archiving data.
		Archiving Sensitive Data The UTL RDM Guide contains information on how to determine if you are dealing with sensitive data when preparing your data for long term storage in an archive.

[6] More information about Creative Commons licensing options is available at: https://creativecommons.org/