Tri-Agency Policy Update

Chantal Ripp
Interim Research Data Management Librarian
April 7, 2021
Overview

- Update on Tri-Agency RDM Policy
- Portage DMP Assistant
- Resources
- Question period
Politique des trois organismes sur la GDR

Politique GDR lancée le 15 mars 2021

Énoncé de la politique :

Stratégies institutionnelles [Établissement postsecondaire]

Plans de gestion des données [Les titulaires de subvention]

Dépôt des données [Les titulaires de subvention]
DMPs

• Speaks to the management of data both during the active phases of your research project and plans for after completion

• To address issues related to data management prior to starting your research projects!

• Tools:
  • While not required for the purposes of this policy, when developing a DMP, researchers are encouraged to consider using the Portage Network’s DMP Assistant [Source: Policy FAQ]
  • There are various other online tools that guide researchers through the elements of a DMP. Researchers can consult discipline-specific examples from organizations like the Digital Curation Centre or the California Digital Library

Tri-Agency Research Data Management Policy

3.2 Data management plans

All DMPs should describe:
- how data will be collected, documented, formatted, protected and preserved;
- how existing datasets will be used and what new data will be created over the course of the research project;
- whether and how data will be shared; and
- where data will be deposited.
# Implementation dates of Tri-Agency RDM Policy

<table>
<thead>
<tr>
<th><strong>Institutional strategies</strong></th>
<th>By March 1, 2023, research institutions subject to this requirement must post their RDM strategies and notify the agencies when they have been completed.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data management plans</strong></td>
<td>By Spring 2022, the agencies will identify the initial set of funding opportunities subject to the DMP requirement. The agencies will pilot the DMP requirement in targeted funding opportunities before this date.</td>
</tr>
</tbody>
</table>
| **Data deposit**             | After reviewing the institutional strategies and in line with the readiness of the Canadian research community, the agencies will phase in the deposit requirement.  
  • **CIHR-funded researchers:** Since January 1, 2008, recipients of CIHR funding have had to comply with the limited data deposit requirements included in the Tri-Agency Open Access Policy on Publications. They must continue to comply with these requirements, which are specific to bioinformatics, atomic, and molecular coordinate data. |
Assistant PGD

Un outil canadien bilingue en ligne qui vous guide tout au long du processus de création un plan !

3 mars, 2021, Assistant PGD 2.0 maintenant disponible
New Portage DMP templates

- Advanced Research Computing
- Arts-Based Research
- Interdisciplinary Health Software/Technology Development
- Mixed Methods (Surveys & Qualitative Research)
- Qualitative Health Sciences Research

- Le développement interdisciplinaire de logiciel ou de technologie en santé
- L’informatique de recherche avancée
- Les méthodes mixtes (enquêtes et recherche qualitative)
- La recherche basée sur les arts
- La recherche qualitative en sciences de la santé

https://portagenetwork.ca/tools-and-resources/training-resources/
More coming...

- CRDCN: Accessing Data from Research Data Centres
- CRDCN: Research Data Centres and External Analysis
- History and the Humanities
- Neuroimaging in the Neurosciences
- Open Science Workflows
- Studying Molecular Interactions
- Systematic Reviews
- Water Quality Research

- RCCDR : L'accès aux données des centres de données de recherche
- RCCDR : Les centres de données de recherche et les analyses externes
- La recherche en histoire et en humanités
- Les études en neuro-imagerie dans les neurosciences
- Les flux de travail en science ouverte
- L’étude des interactions moléculaires
- Les revues systématiques
- La recherche sur la qualité de l’eau

https://portagenetwork.ca/tools-and-resources/training-resources/
NEW templates can be used to create new plans with Portage Network selected as organisation

Create a new plan

Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

* What research project are you planning?

Test DMP

[ ] mock project for testing, practice, or educational purposes

* Select the primary research organisation

Portage Network

[ ] No research organisation associated with this plan or my research organisation is not listed

Which DMP template would you like to use?

Portage Template

Portage Template

Portage Template for Advanced Research Computing

Portage Template for Arts-Based Research

Portage Template for Interdisciplinary Health Software/Technology Development

Portage Template for Mixed Methods (Surveys & Qualitative Research)

Portage Template for Qualitative Health Sciences Research
Version du modèle 0, publiée le 03 mar 2021

Consignes

Ce modèle de plan de gestion de données à méthodologies mixtes a été élaboré pour être utilisé avec l’Assistant PGD de Portage et est destiné à aider les chercheurs qui mènent des recherches à méthodologies mixtes (par exemple, des enquêtes et des entrevues qualitatives ou des groupes de discussion) à élaborer des plans de gestion de données de haute qualité pour soutenir leurs recherches. Le modèle comprend un certain nombre de catégories, de questions et de conseils personnalisés qui sont directement liés à la recherche utilisant des méthodologies mixtes.

Politiques en matière de gestion des données de recherche

- Existe-t-il des politiques en matière de gestion des données de recherche qui définissent les exigences ou les meilleures pratiques concernant la gestion de vos données ? Le cas échéant, indiquez des détails et des liens URL vers ces politiques si nécessaire.

Collecte de données

- Décrivez les types de données que vous allez collecter, y compris toutes les données des enquêtes, des entrevues ou des groupes de discussion. Si d’autres types de données sont recueillis ou générés, décrivez-les également.
- Existe-t-il des données que vous pouvez réutiliser qui répondront à l’une de vos questions de recherche ? Si oui, veuillez expliquer comment vous obtiendrez ces données et les integrirez dans votre projet de recherche.
- Il est important de spécifier et de comprendre le plus tôt possible les méthodes que vous utiliserez pour collecter vos données afin de vous assurer qu’elles répondent à vos besoins, y compris la collecte sécurisée de données sensibles, le cas échéant.

 Créz les méthodes que vous utiliserez pour collecter vos données.

- Si des enregistrements audio d’entrevues ou de groupes de discussion seront transcrits, décrivez comment cette transcription se fera en toute sécurité, notamment si la transcription sera effectuée par l’équipe de recherche ou en dehors de celle-ci (sous-traitée), ou bien si un logiciel, des plateformes ou des services électroniques seront utilisés pour la transcription.
- Décrivez comment vos données seront transférées en toute sécurité, notamment à partir de dispositifs ou de plateformes de collecte de données et à destination ou en provenance des transcrivants, le cas échéant.
- Décrivez tous les formats de fichiers dans lesquels vos données existeront, y compris pour les différentes versions des données d’enquête et d’entrevue qualitative ou de groupe de discussion. Ces formats permettront-ils la réutilisation, le partage et l’accès à long terme aux données?

Documentation et métadonnées

- Décrivez toutes les documentations et métadonnées qui seront utilisées afin de garantir que les données puissent être lues et comprises pendant les phases actives du projet et à l’avenir.
- Décrivez les conventions de nomenclature des fichiers qui seront utilisées afin de soutenir l’assurance qualité et la gestion des versions de vos fichiers et d’aider les autres à comprendre comment vos données sont organisées.
- Décrivez comment vous ferez en sorte que la documentation et les métadonnées soient créées, saisies et, si nécessaire, mises à jour de manière cohérente tout au long du projet de recherche.
- Décrivez les normes et les outils de métadonnées que vous utiliserez pour décrire et documenter vos données.
Storage, Access, and Backup
- 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.
- Describe how members of the research team will securely access and work with data during the active phases of the research project.
- 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.

Preservation
- 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.
- 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.

Sharing and Reuse
- Describe what data you will be sharing, including which version(s) (e.g., raw, processed, analyzed) and in what format(s).
- 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.
- What type of end-user license will you include with your data?

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

Ethics and Legal Compliance
- If applicable, what strategies will you undertake to address secondary uses of data, and especially those which are sensitive in nature?
- How will you manage legal, ethical, and intellectual property issues?
Portage DMP exemplars

- Digital Humanities
  - Data Management Plan for Belgians and French in the Prairies
  - Data Management Plan: Soundscape Study
  - Data Management Plan for Women’s Print History Project (1750-1830)
- Digital Humanities and Secondary Data
  - Data Management Plan for Historical Canadian Census Data
- Mixed Methods
  - Data Management Plan for Mixed Methods Fictional Exemplar
- Natural Sciences
  - Data Management Plan for Ecohydrology Research Group
  - Data Management Plan for Computational reproducibility in High-Performance Computing (HPC)
- Social Sciences
  - Data Management Plan for People, Places, Policies and Prospects: Affordable Rental Housing for Those in Greatest Need
  - Data Management Plan for Usage of Academic Profile Websites

Data Collection

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

We will be collecting surveys which will then be exported into tabular format.

We will also be conducting both semi-structured interviews and focus groups that will produce both digital audio and text (transcriptions) based data.

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?

Our file formats will exist both in non-proprietary and proprietary formats. The non-proprietary formats will ensure that these data are able to be used by anyone wishing to do so once they are deposited and made openly available.

Surveys will exist in .csv (non-proprietary), MS Excel, & SPSS (both proprietary) formats. For more information regarding SPSS see: SPSS.Wikipedia.https://en.wikipedia.org/wiki/SPSS.

Interviews & focus groups data will exist in .mp3 (non-proprietary), MS Word & NVivo (both proprietary) formats. For more information regarding NVivo see: NVivo.Wikipedia.https://en.wikipedia.org/wiki/NVivo

Any survey data deposited for sharing and long-term access will be in .csv format so that anyone can use them without requiring proprietary software.

The final de-identified versions of the interviews and focus group transcripts will be exported into a basic non-proprietary text format for deposit, long-term preservation and access.

If data are collected using laptops or mobile devices, please explain how you will securely store and transfer the data.

Laptops are not being used for any data collection, though encrypted digital voice recorders (DVRs) will be used to collect both interviews and transcripts. Interviews and focus group digital audio files will not be stored on the DVRs, only collected and then securely transferred to the project's cloud based virtual research environment space via a secure FTP (File Transfer Protocol).
Storage and Backup

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

All data storage and back-up procedures will be clearly outlined within the project’s data collection policies and procedures which will be developed prior to data collection.

Survey data will be collected using REDCap, a secure data collection and management software hosted by the Women & Children’s Health Research Institute (WCHRI) at the University of Alberta. REDCap servers undergo regular (daily, weekly, monthly) backups. Secure transferring of data from the REDCap platform to our virtual cloud-based research environment will occur using both encryption and a secure File Transfer Protocol platform. Upon being exported from REDCap, survey data will be immediately ingested into our research project space located on Compute Canada’s cloud platform and which has regularly scheduled backup processes in place.

Qualitative interviews will be conducted using encrypted digital voice recorders. Upon completion of interviews they will be securely transferred within 48 hours to the virtual research project space located on Compute Canada’s cloud platform. Once the audio interviews are uploaded to the cloud platform, they will be deleted from the digital voice recorders.

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

All data will be securely stored in our virtual project space located on Compute Canada’s cloud platform and these will be accessible only by approved researchers, trainees, and study staff. Access to the platform is securely password protected, with access rights ultimately approved by the Principal Investigator and assigned by Compute Canada.

For more information regarding this service provided by Compute Canada see: https://www.computecanada.ca/research-portal/accessing-resources/rapid-access-service/

Preservation

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).

All data will be maintained for a minimum 5 years after the completion of the project, as per University of Alberta ethics requirements.

For the long term, we will be preserving both the raw and the master (cleaned and processed) versions of the surveys. The de-identified and processed versions of surveys will be deposited for long-term preservation and open access. We will be preserving only the de-identified interview transcripts, and these too will be deposited for long-term preservation and open access.

At the end of your research project, where will you deposit your data for long-term preservation and access?

At this time, we believe that we will be using the University of Alberta’s institutional data repository, Dataverse, to deposit and support the long-term preservation, discovery, and access of our data. Dataverse is freely available for use by our project and contains a number of desirable features including the assignment of unique and persistent digital object identifiers (DOIs), the ability to restrict access to data at the file level, built in data citations, data usage metrics, and file versioning. More information regarding Dataverse is available at: https://en.wikipedia.org/wiki/Dataverse

Should the UoA Dataverse be determined to not be able to support any of our data deposit needs, we will be consulting with the UoA Library data team to help guide us through the deposit process in order to find the optimal solutions and support for our project.
For more information on DMPs

DATA MANAGEMENT PLANS

This brief guide provides basic information about data management plans (DMPs). More resources for DMPs can be found on the Portage website, including a Brief Guide for creating an effective DMP.

WHAT IS A DATA MANAGEMENT PLAN (DMP)?

A DMP is a formal document that details the strategies and tools you will implement to effectively manage your data both during your research project and after its completion.

WHY CREATE A DMP?

- Efficiency: identify both strategies and potential challenges in advance; develop sound data practices for your research team; prepare data for effective use during your project.
- Research Quality: ensure reliability and accuracy of data through careful documentation of your data collection, handling, and stewardship practices.
- Reusability and Impact: improve discoverability, accessibility, and reusability of your data by planning for sharing in a repository; increase the potential impact of your research.
- Compliance: Satisfy DMP requirements that may be set forth by specific granting agencies or your own institution.

COMPONENTS OF A DMP

- Data collection: data types, file formats, naming, and version control
- Documentation: ensure data can be read and interpreted
- Data storage and backup throughout the research
- Data preservation strategy for long-term access
- Provisions for sharing and reuse
- Data management responsibilities and resources
- Ethical and legal compliance

CREATE AN EFFECTIVE DATA MANAGEMENT PLAN

This brief guide presents a general framework for creating an effective data management plan (DMP) to help you plan and organize your research and to meet research funder requirements.

To prepare your DMP, visit the Portage DMP Assistant tool.

GENERAL GUIDELINES

- Begin by providing a description of your research project, its focus, and purpose.
- Avoid the excessive use of discipline-specific jargon - your DMP should be easily understandable by anyone.
- Provide clear rules for any acronyms used.
- Do not leave sections or questions blank.
- Provide rationale for decisions made - help others understand why you made a decision.
- Your DMP is a living document - update it as needed!

DATA COLLECTION

- Include descriptions of how you will collect data, including from where and in what format(s).
- Provide an estimate of the amount of data you will collect (e.g., MiB, GiB, TBs).
- Describe any software or platforms that will be used for data collection.
- Clearly explain how you will store and transfer data.
- Establish how you will organize your data, including details relating both to file naming and versioning.

DOCUMENTATION AND METADATA

- Describe what information will be needed for others to understand or reuse your data.
- Describe how you will consistently capture documentation throughout the project.
- Choose a metadata standard suited to your discipline and/or chosen data repository or provide rationale for creating your own.

DATA MANAGEMENT PLANS

What is a data management plan (DMP)?

A DMP is a formal document that details the strategies and tools you will implement to effectively manage your data during the active phase of your research, and the mechanisms you will use for preserving and appropriately sharing your data at the end of the project. A DMP is a "living" document that can be modified throughout your project to reflect any changes that have occurred. More resources for DMPs can be found on the Portage website, including a Brief Guide for creating an effective DMP.

Why create a DMP?

- A DMP helps you:
  - Meet grant application requirements and/or adhere to institutional data management plans.
  - Make it easier for all team members to document, understand, find, and use the data.
  - Plan the resources, tools, and expertise needed for data management.
  - Identify challenges for storing, handling, and managing the types and volume of data.
  - Ensure reliability, authenticity, accuracy, and irreducibility of your data.
  - Have a detailed account of your data collection, handling, and stewardship practices.
  - Plan how to make your data FAIR (findable, accessible, interoperable, and reusable) to maximize the research potential and impact of your data.

Standard Components of a DMP

Data collection:

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

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 and accessible at the end of your project.

http://doi.org/10.5281/zenodo.4495482
http://doi.org/10.5281/zenodo.4004957
http://doi.org/10.5281/zenodo.4495631
Data repositories for archiving and preservation

Repositories provide:
• Technical infrastructure
• Visibility for your research data
• Persistent identifier (DOI) and citation for your data
• Access controls
Choosing a data repository

- Recommendations or requirements in funder guidelines (e.g. CIHR Open Access Policy: Annex)
- Recommended repositories maintained by journals/publishers (e.g. Scientific Data and PLOS ONE)
- Consult library website for supported repositories, including uOttawa Dataverse and Federated Research Data Repository (FRDR)
- Consult FAIRsharing or re3data.org – provide a list of certified data repositories
Generalist repositories

- Scholars Portal Dataverse
- Zenodo
- Open Science Framework
- Federated Research Data Repository (FRDR)
- Figshare
Repository – Scholars Portal Dataverse
Deposit dataset(s) and documentation

- data file(s)

- Provide documentation and supporting information (e.g. codebook and/or Readme file; methodologies; data provenance).

- Portage’s Documentation and Supporting Material Required for Deposit
Questions?

Chantal Ripp
Research Data Management Librarian (interim)
Chantal.ripp@uottawa.ca

Guide de GDR à Université d'Ottawa en français

RDM at uOttawa guide in English