McMaster University

Family Medicine

David Braley Primary Care Research Collaborative

@McMasterFamMed

# Data Management for Tri-Agency Grants

Watch a recording of the session here

**Rebecca Clark**

January 17th, 2023

Michael G. DeGroote SCHOOL OF MEDICINE

McMaster University

Family Medicine

# Data Management for Tri-Agency Grants

RKSB January 17th, 2023

Presented by: Rebecca Clark

@McMasterFamMed

Michael G. DeGroote SCHOOL OF MEDICINE

# Objectives of RKSB Session

1. Provide an overview of the Tri-agency grant research data management (RDM) requirements
2. Describe how McMaster is currently/plans to support these requirements
3. Share resources to help applicants prepare for and meet requirements

# Tri- Agency Requirements for Grant Applications

1. Institutional RDM Strategy (*March 2023*)
2. Deposit data in a repository *(Estimated Fall 2024)*
3. All grant applications will need a data management plan (DMP) at submission
   - Pilot (*Spring 2023/Fall 2023)*
   - All grants (*Estimated Fall 2024)*

Note: The information provided is what is known as of January 2023.

# 1. Institutional Research Data Management Strategy

# Institutional Research Data Management Strategy

- Posted publicly
- High level description
- Living document
- Will differ based on institution
  - e.g. size, existing capacity to support RDM
- Strategies won't be evaluated by Tri-Agency

# Institutional Research Data Management Strategy

- Describes how McMaster will create an environment that supports and enables RDM
- Sections include:
  - Organizational frameworks and support
  - RDM practices, tools and infrastructure
  - Indigenous data
  - Timeline of activities

https://rdm.mcmaster.ca/rdm-strategy

# 2. Data Repositories

# TRUE OR FALSE

**The Tri-Agency requires all data deposited in a repository to be open access?**

# False

# Data Repositories

- Can choose which repository and level of access
  - Open access is encouraged
- Data deposit may looking different depending on data type
  - E.g., PHI or Qualitative data
- All data, metadata and code that relate to publications/pre-prints should be deposited
- Deposit required by the time of publication

# MacSphere - Institutional

- Must be associated with McMaster
- Resources to support deposit
- Set level of access
- Available for different data types
- Not recommended for sensitive data

https://macsphere.mcmaster.ca/

# McMaster University Dataverse - National

- Canadian Dataverse Repository (Borealis)
- Supported by many institutions across Canada
- Hosted by Scholars Portal and University of Toronto Libraries (Canada)

borealis The Canadian Dataverse Repository
Le dépôt Dataverse canadien

# McMaster University Dataverse - National

## Features

- Generates DOI
- Templates available
- Metrics application
- Dataset and version tracking
- Flexible permissions and levels of access
- Has test site to practice uploading/see how it all works
- And more…

https://borealisdata.ca/dataverse/mcmaster

# Federated Research Data Repository

- Holds large (1 TB+) data sets
- Free for individuals with a Canadian institution
- Dedicated staff to assist with deposits and metadata
- Ultimately open access
- Pilot for sensitive data
  - Optional zero-knowledge encryption

https://www.frdr-dfdr.ca/repo/?locale=en

# Repositories Finders

- NIH Scientific Data Sharing: NIH supported repositories

https://sharing.nih.gov/accessing-data/accessing-scientific-data

- FRDR: Database of Canadian repositories

https://www.frdr-dfdr.ca/discover/html/repository-list.html?lang=en

- Data Cite

https://repositoryfinder.datacite.org/

- McMaster RDM Data Storage Finder

https://rdm.mcmaster.ca/finder

# 3. Data Management Plans

# TRUE OR FALSE

**Other than grant applications, DMPs are a waste of time**

# FALSE

# Data Management Plans

- Part of good data management practices
  - Helps with integrity, usability, (and re-using), preservation, etc
- Describes the practices, processes and strategies related to data management
- Proactively think about data management (including challenges that might come up)
- Should be modified throughout the research process

# Tri-Agency Requirements for DMPs

DMPs should describe:

1. How data will be **collected, documented, formatted**, **protected and preserved** over time
2. Existing datasets being used and new data that will be created throughout the study
3. How and if data from the study will be **shared**
4. Where the data will be **deposited**

# Tri-Agency Requirements for DMPs

5. Who is **responsible** for the data
6. **Roles** and responsibilities of team members
7. **Ethics, legal and commercial** constraints of the data
8. If involving First Nations, Metis, and Inuit communities, collectives and organizations
   - Co-developed and recognize Indigenous data sovereignty
   - OCAP
   - CARE

# Tri-Agency Requirements for DMPs

- More details will be provided based on pilot phase
- Approximately 2-3 pages
- Not scored in pilot phase, but will be scored with full roll out
- Do not have to use a DMP assistant

# DMP Assistant

# DMP Assistants

## Portage

- Free
- Bi-lingual
- Hosted in Canada
- Resources available
- Can share with collaborators
- Can export plan

https://assistant.portagenetwork.ca/

# My plan (Portage Template)

Project Details | Contributors | Plan overview | Write Plan | Share | Request feedback | Download

expand all | collapse all

5/20

- Data Collection (3 / 3) +
- Documentation and Metadata (2 / 3) +
- Storage and Backup (0 / 3) +
- Preservation (0 / 2) +
- Sharing and Reuse (0 / 3) +
- Responsibilities and Resources (0 / 3) +
- Ethics and Legal Compliance (0 / 3) +

# TRUE OR FALSE

# Information in DMPs are things we already think about in research?

# TRUE

Data in DMPs is very similar to ethic applications

OR

Information/processes that are common parts of research and grant processes

There is very little (if any) new information that you need to think about for DMPs

# DMP – Data Collection

- **What types of data will you collect, create, link to, acquire and/or record?**
  - 6.1 Study Protocol (measurements and analysis)
  - 6.7/6.8  What is your your primary outcome/secondary outcomes and how will it be measured
- **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?**
  - 6.1 Study Protocol (measurements and analysis)
  - New? Related to data preservation

# DMP – Data Collection

- **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?**
- Typically part of research processes and folder/file set up

# DMP – Documentation and Metadata

- **What documentation will be needed for the data to be read and interpreted correctly in the future?**
  - Typically part of research processes
  - Examples: Protocol, Data collection forms, Data dictionaries, Read me files

- **How will you make sure that documentation is created or captured consistently throughout your project?**
  - Typically part of research process

# DMP – Documentation and Metadata

- **If you are using a metadata standard and/or tools to document and describe your data, please list here.**
- New? May not apply to all applications.
- Related to data preservation and what data to include within your metadata

# DMP – Storage and Backup

- **What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?**
  - New? Can refer to past/existing projects or consult IT
  - 14.5 How long will you keep the study data
  - Some funders require certain lengths of time

- **How and where will your data be stored and backed up during your research project?**
  - 14.2 How will you store and protect the study code (or other data with identifiers)

# DMP – Storage and Backup

- **How will the research team and other collaborators access, modify, and contribute data throughout the project?**
  - May not think about it, but part of research process already
  - Example: If storing data on Sharepoint, collaborators can securely access data.

# DMP - Preservation

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

- **Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.**
  - New? Touched on this previously
  - Section 12: Personal Health Information

# DMP – Data Preservation Examples

- Tabular data: CSV
- Structured plain text: XML, JSON
- Images: PDF, JPEG, PNG, TIFF
- Audio: FLAC, MP3, AIFF
- Video: MOV, MPEG-4, MKV
- Compressed files: ZIP, TAR.GZ

# DMP – Sharing and Reuse

- **What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).**
  - Many publications and funders already ask for data sharing statements

- **Have you considered what type of end-user license to include with your data?**
  - New?
  - Examples: Creative commons licenses, open data commons licenses

# DMP – Sharing and Reuse

- **What steps will be taken to help the research community know that your data exists?**
  - New? Typically part of research Knowledge translation strategies

# DMP –Responsibilities and Resources

- **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.**
  - Section 2: Investigator and Training Information
  - Part of grant application/budget.
- **What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?**
  - Part of grant application/budget
  - 15.1/15.2 Budget

# DMP –Responsibilities and Resources

- **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?**
  - New?

# DMP – 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?**
  - Ethics: Related to data collection, data storage, and sercurity
  - Section 12: Personal Information
    - Personal Health Information, Storage and Technical requirements

- **How will you manage legal, ethical, and intellectual property issues?**
  - Related to data sharing, licensing, ownership of data

- **If applicable, what strategies will you undertake to address secondary uses of sensitive data?**
  - Ethics: covered in ethics applications and consent processes
  - Repository access

# DMP – Ethics and Legal Compliance

## Information for HiREB and DMPs

- **Encryption**: All user communication with the McMaster University – DFM SharePoint document repository across the Internet uses SSL / TTL connections and all SSL connections are established using 2048-bit keys encryption
- **Firewalls**: DFM uses BitLocker. The BitLocker is a full volume encryption feature included with Microsoft Windows.  We use University enterprise level firewall as well, SonicWall NSA series appliance to protect our networks.
- **Anti-virus software**: Microsoft Defender Antivirus

# Exemplar a Mixed Method DMP

https://zenodo.org/record/4019563#.Y8Vv1HbMI2w

# 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 groups transcripts will be exported into a basic non-proprietary text format for deposit, long-term preservation and access.

**How much data do you anticipate collecting? Include an estimate of how much storage space you will require (in megabytes, gigabytes, terabytes). This estimate should also take into account storage space required for file versioning, backups, and the growth rate over time.**

We estimate that we will be collecting approximately 800 surveys, 20 interviews (approximately 30 min in length each), and 2 focus groups (approximately 90 min in length each). Total magnitude of data, including accounting for versions (raw, master, analytic) is estimated to be under 30GB.

## 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/

# 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.

# 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.**

We plan to share processed survey data. This will include responses at the participant level, with all direct identifiers removed, and including derived variables used for analytic purposes.

Regarding the qualitative interviews, we plan to share processed and de-identified transcripts.

We will obtain participant consent for sharing these data and will additionally be depositing and making available an example of both our study information letter and participant consent form, as well as our approved institutional ethics application.

# Resources

# Resources from DFM

- RKSB on data management (including Portage demonstration)

Video: https://www.youtube.com/watch?v=fHA0kk59AEM

Slides: https://medm-backend.fammedmcmaster.ca/wp-content/uploads/2020/12/Managing-your-research-data-Intro-to-the-DMP-Assistant-slides.pdf

- PCRC Data Management SOP

https://pcrc.fammedmcmaster.ca/wp-content/uploads/2021/10/PCRC_Research-Data-Management-SOP-2021-09-07-1.pdf

# Resources at McMaster

- Research Data Management Services
  - Consultations on data management, DMPs, and more
  - Assist with data sharing/repositories
  - Resources about security, DMPs, organizing data, storing data and publishing data
  - Information about the Institutional strategy (https://rdm.mcmaster.ca/rdm-strategy)

- Sherman Centre for Digital Scholarship
  - Workshops
  - Consults

# Upcoming Workshops at McMaster

- Best Practices for Managing Data in your Research
  - Jan 18th 10:30-11:30am
- Qualitative Data: Practices for RDM Planning and Sharing
  - April 5th 10:30-11:30am
- Depositing & Sharing Data Online with McMaster Dataverse
  - Jan 25th 10:30-11:30am

Workshops are recorded and posted. https://rdm.mcmaster.ca/events

# Other Resources

- Digital Research Alliance of Canada (https://alliancecan.ca/en)
  - Examples of DMPs (https://alliancecan.ca/en/services/research-data-management/learning-and-training/training-resources#heading-dmp-exemplars)
- Many universities have resources/guides/examples available
- OCAP Principles (https://fnigc.ca/ocap-training/)
- CARE Principles (https://www.gida-global.org/care)

# Questions?

McMaster University

# Family Medicine

Department of Family Medicine
Michael G. DeGroote School of Medicine
Faculty of Health Sciences

fammedmcmaster.ca
@McMasterFamMed