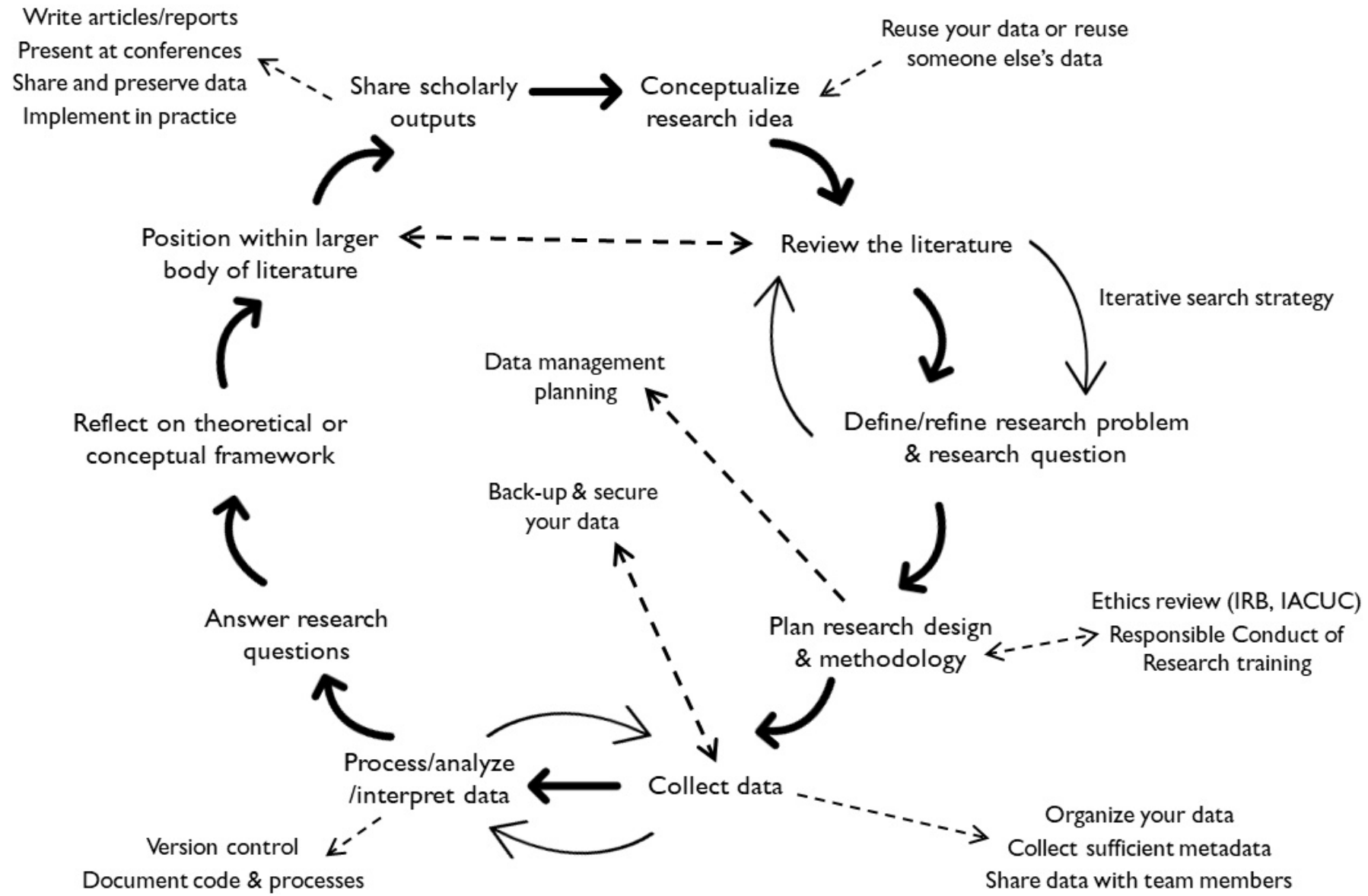
Data Management Plans

Elizabeth Stregger

DATA 3101

October 4, 2022



Copyright and related rights for this image waived via CC0.

Tri-Agency Research Data Management Policy

Canadian Institutes of Health Research (CIHR), Natural Sciences and Engineering Research Council of Canada (NSERC),Social Sciences and Humanities Research Council of Canada (SSHRC) form the Tri-Agency

Tri-Agency Research Data Management (RDM) Policy (2021)

- Institutional Strategy
- Data Management Plans
- Data Deposit
- https://www.science.gc.ca/eic/site/063.nsf/eng/h_547652FB.html

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.

(Portage, 2021)

See: 10.5281/zenodo.4015610

Some of the examples you chose for the Retraction Watch assignment included problems after completion (not preserving raw data or code used in analysis).

Why? Efficiency, Research Quality, Reusability and Impact, Compliance (Portage, 2021)

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

Consider yourself & your team

- Data management planning will save you time at the end of your project
- Discuss questions of data ownership and policies early to prevent miscommunication and conflict
- Focus on the best practices that are most important for your project:
  - Data about humans: security and legal and ethical considerations
  - Data from instruments: file structure and data dictionaries
  - Collaborations: ownership and data roles
- Use tools (the DMP Assistant) and ask for help!

My data management plan – a satire

“I will store all data on at least one, and possibly up to 50, hard drives in my lab. The directory structure will be custom, not self-explanatory, and in no way documented or described. Students working with the data will be encouraged to make their own copies and modify them as they please, in order to ensure that no one can ever figure out what the actual real raw data is.

Backups will rarely, if ever, be done.”

(Brown, 2010)

Read more at: http://ivory.idyll.org/blog/data-management.html

Data Management Plan as an active document

- A Data Management Plan might look great to the funder, but if it isn't followed, it was a waste of time
- Changes to your research plan or roles should be updated in the DMP
- Version control your data management plan in the same way you'd version control other documents in your research (protocols, scripts, etc.)
- Recommend: reviewing your DMP as a group every 6 months
- Share the updated Data Management Plan and documents with the whole team

Portage Templates

- Accessing Data from Research Data Centres (StatCan)
- Advanced Research Computing
- Arts Based Research
- Interdisciplinary Health Software / Technology Development
- Mixed Methods (Surveys & Qualitative Research)
- Neuroimaging
- Open Science Workflows
- Qualitative Health Sciences Research

https://assistant.portagenetwork.ca/public_templates

Check-in

Which template is the best match for the data you found for the Find Data and/or OpenRefine assignment?

In-class example: Tick Data

Together, we'll review the template for Open Science Workflows with the Tick Data example in mind

Reminder: this data set is published on https://open.canada.ca/en/open-data

Link: https://open.canada.ca/data/en/dataset/a3994d9e-ed40-fa6a-372d-f9820eae269d

I'll create a first draft of a Data Management Plan for this project

Looking at the tick data example

- How was it described?
  - 2012-2018
  - Research at Mount Allison
  - Published on government of NB open data portal, with a map

Need more information on: what data was submitted by community members, how identification of ticks and testing for Borrelia was done in the lab, and how the geospatial data was assigned.

Tick data example: file formats

- What file formats are included in the data deposit?
  - CSV data table – enter location data from community members, entered results of analysis
- Need to get from there to mapping – what software am I using? How will I go from a description of a location to specific lat/long for graphing? Look at file types:
  - Kml files – probably originated in Google Earth, can be opened in ArcGIS and QGIS
  - GeoJSON – open file type, can be easily viewed in QGIS
  - Shp – best in ArcGIS, can be viewed in QGIS

Assignment

- Read the data management plan template that is most closely aligned with the topic of the data you chose for the Find Data and/or OpenRefine assignment
- Create an account on the DMP Assistant website: https://assistant.portagenetwork.ca/
- Pretending you are the researcher, create a data management plan for this project
- Download the file, put it in your personal repository on your computer, and push it to GitHub. Create an issue to let me know it is ready for feedback.

Reminder

- Mid-term assessment check-ins will be scheduled for October 25 & 27
- If you haven't already finished some assignments and created issues to notify us, please do!
- For ungrading to work, we need to be in communication.

What is preventing you from submitting? Can I help? (Asking as a former procrastinating perfectionist…)