Qualitative Data: Practices for RDM Planning and Sharing

Danica Evering, MA

April 5, 2023

Research Data Management Services

McMaster University

McMaster University | Library

McMaster University is located on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the “Dish With One Spoon” wampum agreement.

Mhsheikholeslami, “Eramosa Karst Conservation Area- Nexus Cave- Stoney Creek- Hamilton-Ontario,” 16 June 2019, Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Eramosa_Karst_Conservation_Area-_Nexus_Cave-_Hamilton-Ontario-2.jpg

Code of Conduct

The Sherman Centre and the McMaster University Library are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

Please refer to our code of conduct webpage for more information:
scds.ca/events/code-of-conduct/

Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences. In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

Certificate Program

The Sherman Centre offers a Certificate of Completion that rewards synchronous participation at 7 workshops. We also offer concentrations in Data Analysis and Visualization, Digital Scholarship, and Research Data Management.

Learn more about the Certificate Program: https://scds.ca/certificate-program
If you would like to be considered for the certificate, verify your participation in this form: https://u.mcmaster.ca/verification

At an unspecified point during the workshop, a code will be read aloud. This is the answer to the third question of the form.

Outline

What does RDM look like with qualitative research?

Planning – qualitative data management plans + ethics

Data Collection + Analysis – security and documentation

Publication – considerations for sharing and archiving

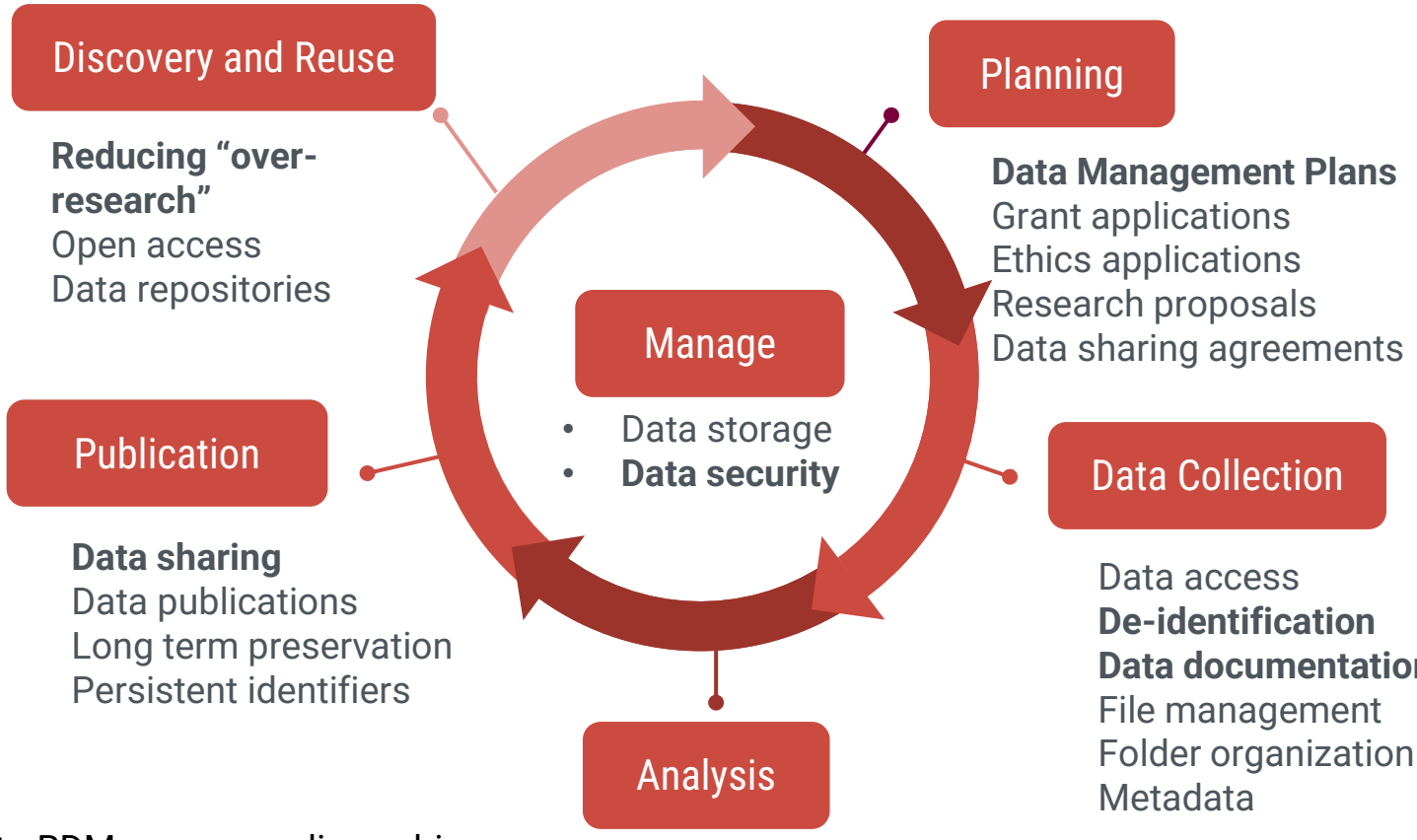
Discussion – what are some limitations and curiosities

"Data may be in any format or medium taking the form of text, numbers, symbols, images, films, video, sound recordings, pictorial reproductions, drawings, designs or other graphical representations, procedural manuals, forms, diagrams, workflows, equipment descriptions, data files, data processing algorithms, software, programming languages, code, or statistical records."

Innovation, Science and Economic Development Canada. "Frequently Asked Questions -Tri-Agency Research Data Management Policy." Government of Canada. Innovation, Science and Economic Development Canada, October 19, 2021. https://science.gc.ca/eic/site/063.nsf/eng/h_97609.html#1a

Research Data Management for Qualitative Data

Research Data Management is the active organization & maintenance of data throughout the research data lifecycle to ensure its **security**, **accessibility**, **usability**, and **integrity**.



For a fuller introduction to RDM see our earlier webinar "Best Practices for Managing Data in your Research" - https://scds.github.io/intro-rdm/intro.html

McMaster University | Library

Photo by LinkedIn Sales Solutions on Unsplash

Why is RDM important?

- Recognize the value of datasets.
- Avoids duplication + increases value for both for research community and funder.
- Reducing research impact on participants.
- Public access means greater transparency and citizen research.

McMaster University | Library

Kevin Patrick Robbins, “Moms to Babies Research, Focus Group,” June 19, 2018, McMaster Asset Bank, https://brand-resources.mcmaster.ca/asset-bank/action/viewAsset?id=7293&index=21&total=68&view=viewSearchItem

Planning – RDM best practices for starting research with qualitative data.

McMaster University | Library

Data Management Plans (DMPs)

- A **Data Management Plan (DMP)** is your plan for how you will create, store, organize, document, secure, preserve, and share your research data.
- A document which speaks to the management of data both **during** the active phases of your research and **after** the completion of the research project.
- Helpful to do this at the same time as an ethics application!

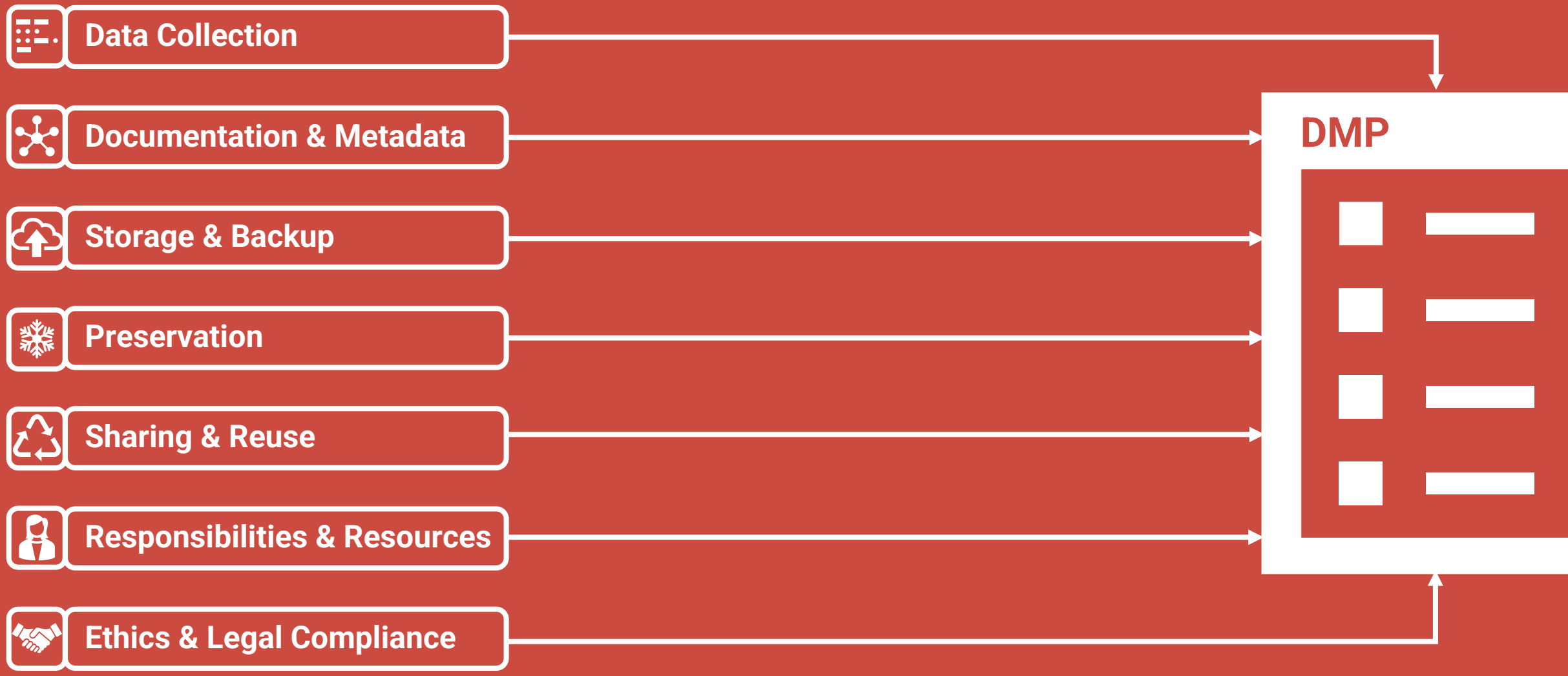
Photo by Hanna Morris on Unsplash.

Why create a DMP for **qualitative data?**

- Blueprint for how you will handle data during your project - avoid pitfalls and problems before they occur.
- Establish systems for secure storage, translation, etc. to protect research participants.
- Prepare for future stages of research including potential data sharing.
- Funders requiring applicants to submit a DMP. Tri-Agencies (NSERC, CIHR, SSHRC – started 2022), NIH (starting 2023), IDRC (2017), and others.

Photo by Soundtrap on Unsplash.

What goes in a Data Management Plan?



RDM Services: **Data Management Plans (DMPs)**

- A web-based, bilingual data management planning tool.
- Available to all researchers in Canada.
- Walks you through relevant questions for data management.
- Templates and exportable plans.
- Send to RDM Services for review!
- Access at **rdm.mcmaster.ca/plan**

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

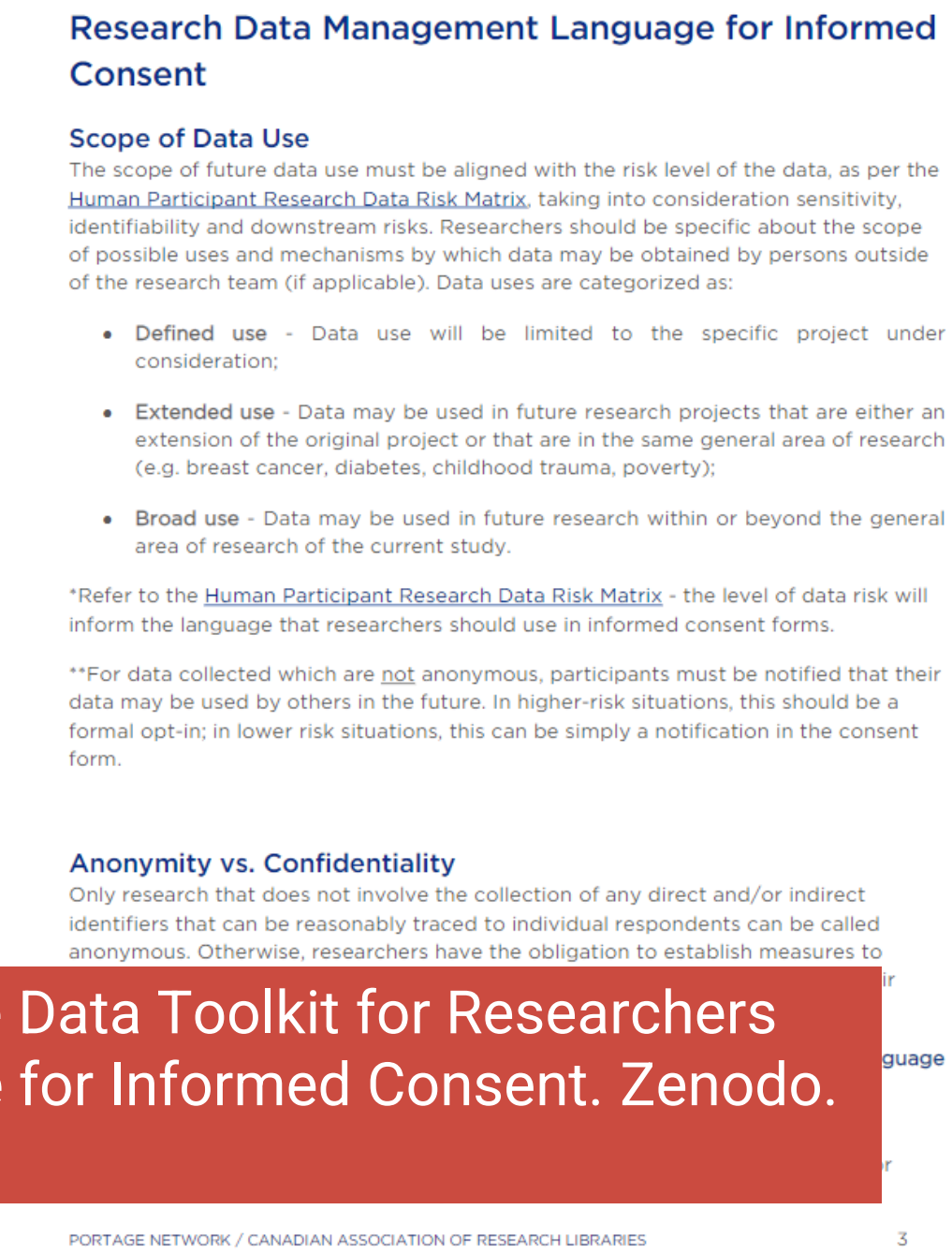
DMP Example: **Mixed Methods**

https://alliancecan.ca/en/services/research-data-management/learning-and-training/training-resources

McMaster University | Library

RDM Language for Informed Consent

- What data may be used for – **defined, extended, broad use**?
- What are their expectations of **anonymity**?
- Who will have **access** to this data and how?
- When and how can they **withdraw** data?
- How can data be used in the **future**?



Research Data Management Language for Informed Consent

Scope of Data Use

The scope of future data use must be aligned with the risk level of the data, as per the Human Participant Research Data Risk Matrix, taking into consideration sensitivity, identifiability and downstream risks. Researchers should be specific about the scope of possible uses and mechanisms by which data may be obtained by persons outside of the research team (if applicable). Data uses are categorized as:

- **Defined use** - Data use will be limited to the specific project under consideration;
- **Extended use** - Data may be used in future research projects that are either an extension of the original project or that are in the same general area of research (e.g. breast cancer, diabetes, childhood trauma, poverty);
- **Broad use** - Data may be used in future research within or beyond the general area of research of the current study.

*Refer to the Human Participant Research Data Risk Matrix - the level of data risk will inform the language that researchers should use in informed consent forms.

**For data collected which are not anonymous, participants must be notified that their data may be used by others in the future. In higher-risk situations, this should be a formal opt-in; in lower risk situations, this can be simply a notification in the consent form.

Anonymity vs. Confidentiality

Only research that does not involve the collection of any direct and/or indirect identifiers that can be reasonably traced to individual respondents can be called anonymous. Otherwise, researchers have the obligation to establish measures to

PORTAGE NETWORK / CANADIAN ASSOCIATION OF RESEARCH LIBRARIES 3

Sensitive Data Expert Group. (2020). Sensitive Data Toolkit for Researchers Part 3: Research Data Management Language for Informed Consent. Zenodo. https://doi.org/10.5281/zenodo.4107178

> The entire data lifecycle in research in Africa - the collection, storage, access, use, retention, and disposal of human bio-specimens in bio-banking - could significantly be touching and affecting subtle indigenous cultural underpinnings and affect how research is often received or even perceived in such contexts.

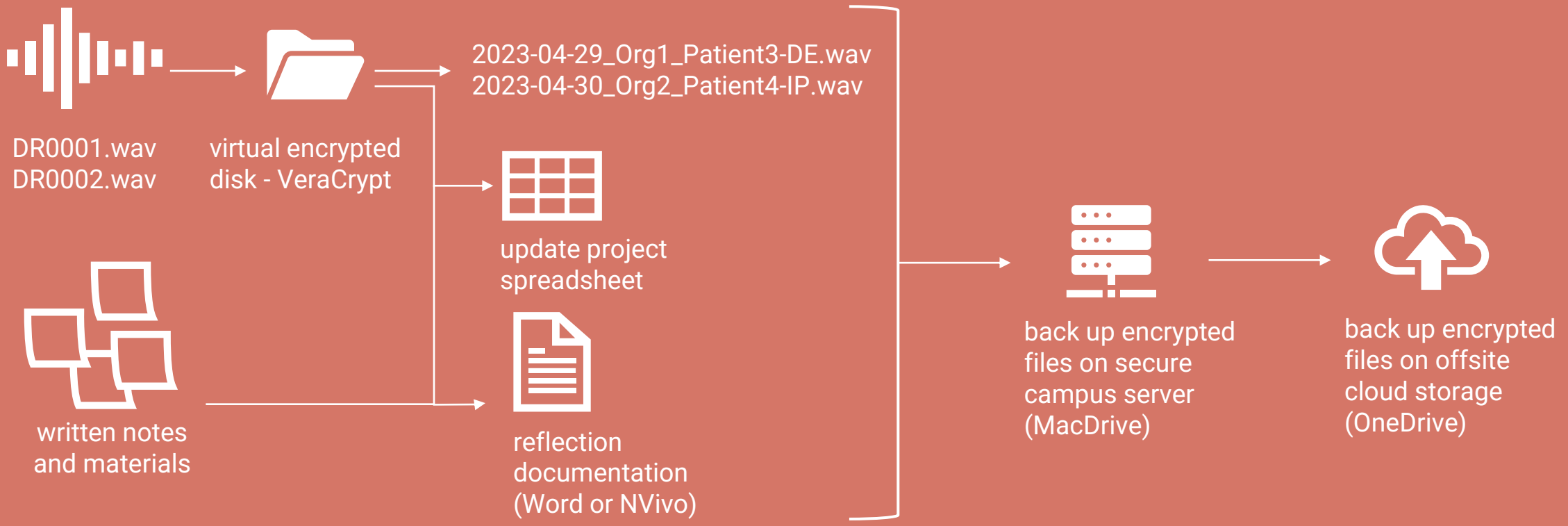
Nderitu, David, and Claudia Emerson. “The Indigenous African Cultural Value of Human Tissues and Implications for Bio-Banking.” *Developing World Bioethics* n/a, no. n/a. Accessed March 31, 2023. https://doi.org/10.1111/dewb.12390.

Data Collection + Analysis – RDM best practices for qualitative data.

McMaster University | Library

Data Collection - **Focus Groups and Interviews**

- **Recording Devices** are better than using your phone! *Can be encrypted*.
- **Space** - background noise, interruptions, quiet voices
- **Testing** and listening back
- **>1 device**, batteries, SD cards



File Storage Workflow – **Focus Group/Interview**

Study Number 6377

Integrated Floodplain Management, 2006-2008

Morris, J.

Floodplain farm survey

Interview ID	Farmer code	Age	Farm scheme	Farm type	Size of farm (hectare)	Number of holdings	Date of interview	Interviewer name	No of pages	Text file name	Audio file name
1	Be1	35-45	Beckingham	Beef	360	1	04.12.2006	Helena	28	6377int001	6377int001
2	Be2	45-55	Beckingham	Arable	364	1	05.12.2006	Helena	21	6377int002	6377int002
3	Be3	45-55	Beckingham	Arable	372	2	06.12.2006	Helena	22	6377int003	6377int003
4	Be4	45-55	Beckingham	Arable	194	3	06.12.2006	Helena	18	6377int004	6377int004
5	Be5	55-65	Beckingham	Arable	108	1	07.12.2007	Helena	21	6377int005	6377int005
6	Be6	45-55	Beckingham	Arable	1254	2	01.02.2008	Helena	19	6377int006	
7	Bu1	55-65	Bushley	Mixed	101	2	13.02.2007	Quentin	29	6377int007	6377int007
8	Bu2	>65	Bushley	Mixed	97	1	15.02.2007	Quentin	15	6377int008	6377int008
9	Bu3	>65	Bushley	Arable	194	4	13.02.2007	Quentin	21	6377int009	6377int009
10	Bu4	55-65	Bushley	Mixed	202	1	15.03.2007	Helena	19	6377int010	6377int010
11	Cu1	35-45	Cuddyarch	Dairy	64	1	08.05.2007	Helena	19	6377int011	6377int011
12	Cu2	55-65	Cuddyarch	Dairy	189	2	08.05.2007	Helena	18	6377int012	6377int012
13	Cu3	55-65	Cuddyarch	Mixed livestock	76	1	08.05.2007	Helena	13	6377int013	6377int013
14								Helena	24	6377int014	6377int014
15								Helena	14	6377int015	6377int015
16								Helena	20	6377int016	6377int016
17								Helena	22	6377int017	6377int017
18								Quentin	17	6377int018	6377int018a
18								Quentin	17	6377int018	6377int018b
19	Id1b	55-65	Idle	Arable	158	3		Quentin	22	6377int019	
20	Id2	45-55	Idle	Dairy	150	1	08.02.2007	Quentin	17	6377int020	6377int020

UK DataService – Example Data List

https://ukdataservice.ac.uk/learning-hub/research-data-management/document-your-data/data-level/data-documentation-qualitative-data/

Transcription

- **Human or AI Transcription**
- **Privacy Policy**: How data can be used + responsibilities for safety.
- **Server Location**: Laws in server location – data privacy laws should be robust.
- Data use as **training model** for AI systems
- Include in **Ethics Review**
- **Guidance doc** coming soon!

"Nowadays they get a tin an' there's nae eggs .in it an' the goodness is oot [out] o' it. Like... have you ever had frozen stuff an' you've cooked it an' you feel as though it didnae taste right... efter ha'in a fresh bit o' steak an' onion. Ken [know] fit [what] I mean? The juice... there's nae the juice in it. Well, that's whit we find wi' the things nowadays an' a, the richt good is out o' them... the body-buildin' material.. afore you eat it. I mean, tinned soup, I would niver hae it in the hoose unless it wis maybe Karen comin' in an' I wis gaun away in a hurry an' gettin' a tin o' soup... I wouldnae gie it to him... we were nae brought up like that, we wis brought up to get a' thing oot o' the groun' and intae a pot..."

Libby Bishop, "A Reflexive Account of Reusing Qualitative Data: Beyond Primary/Secondary Dualism," Article (Sociological Research Online, May 30, 2007), https://www.socresonline.org.uk/12/3/2.html.

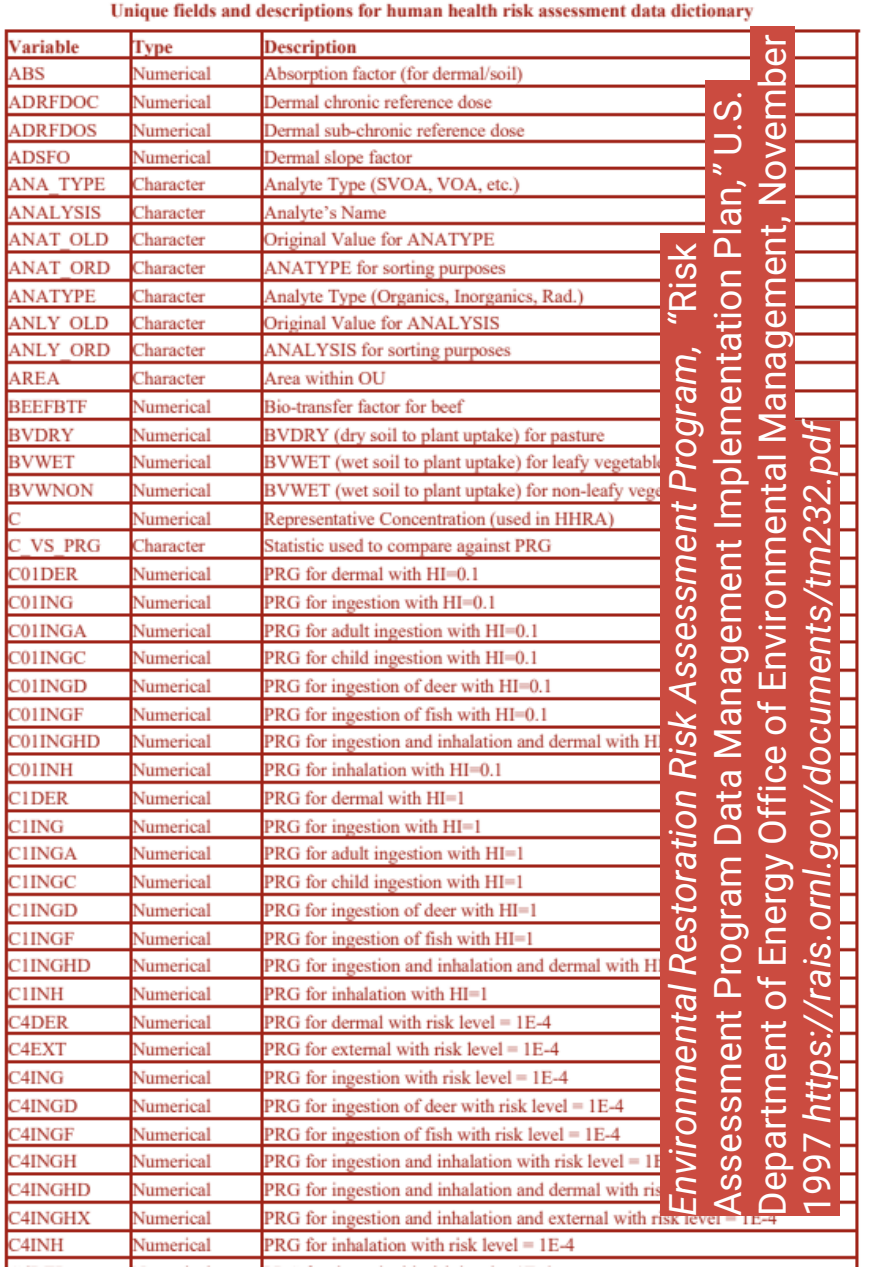
Data Collection - **Surveys**

- **McMaster Surveys:** REDCap, Nvivo, Qualtrics, LimeSurvey
- **Multi-Factor Authentication**: provide greater security for log-ins.
- **Roles and Permissions:** Limit data that each user can see
- **Activity Logs**: Available for QA and data security review
- **Bot Attacks**: CAPTCHA, unique IP restriction, email validation, timestamps, consent consequences

Download Data for Analysis

- Limited analysis functionality in survey software
- **Data Export:** SPSS, SAS, R , STATA
- **De-Identification Options**: Remove identifier fields, invalidated text fields, date shifting, hashing record names before export
- **Codebooks + Data Dictionaries**: Export documentation from survey software
- **Secure Storage**: Ensure secure storage and handling following export.

For more on Sensitive Data Management see our earlier webinar “Securely Managing and Publishing Sensitive Data” - https://scds.github.io/intro-rdm/sensitive.html



Unique fields and descriptions for human health risk assessment data dictionary

Variable	Type	Description
ABS	Numerical	Absorption factor (for dermal/soil)
ADRFDOC	Numerical	Dermal chronic reference dose
ADRFDOS	Numerical	Dermal sub-chronic reference dose
ADSFO	Numerical	Dermal slope factor
ANA_TYPE	Character	Analyte Type (SVOA, VOA, etc.)
ANALYSIS	Character	Analyte's Name
ANAT_OLD	Character	Original Value for ANATYPE
ANAT_ORD	Character	ANATYPE for sorting purposes
ANATYPE	Character	Analyte Type (Organics, Inorganics, Rad.)
ANLY_OLD	Character	Original Value for ANALYSIS
ANLY_ORD	Character	ANALYSIS for sorting purposes
AREA	Character	Area within OU
BEEFBTF	Numerical	Bio-transfer factor for beef
BVDRY	Numerical	BVDRY (dry soil to plant uptake) for pasture
BVWET	Numerical	BVWET (wet soil to plant uptake) for leafy vegetabl
BVWNON	Numerical	BVWET (wet soil to plant uptake) for non-leafy veg
C	Numerical	Representative Concentration (used in HHRA)
C_VS_PRG	Character	Statistic used to compare against PRG
C01DER	Numerical	PRG for dermal with HI=0.1
C01ING	Numerical	PRG for ingestion with HI=0.1
C01INGA	Numerical	PRG for adult ingestion with HI=0.1
C01INGC	Numerical	PRG for child ingestion with HI=0.1
C01INGD	Numerical	PRG for ingestion of deer with HI=0.1
C01INGF	Numerical	PRG for ingestion of fish with HI=0.1
C01INGHD	Numerical	PRG for ingestion and inhalation and dermal with H
C01INH	Numerical	PRG for inhalation with HI=0.1
C1DER	Numerical	PRG for dermal with HI=1
C1ING	Numerical	PRG for ingestion with HI=1
C1INGA	Numerical	PRG for adult ingestion with HI=1
C1INGC	Numerical	PRG for child ingestion with HI=1
C1INGD	Numerical	PRG for ingestion of deer with HI=1
C1INGF	Numerical	PRG for ingestion of fish with HI=1
C1INGHD	Numerical	PRG for ingestion and inhalation and dermal with H
C1INH	Numerical	PRG for inhalation with HI=1
C4DER	Numerical	PRG for dermal with risk level = 1E-4
C4EXT	Numerical	PRG for external with risk level = 1E-4
C4ING	Numerical	PRG for ingestion with risk level = 1E-4
C4INGD	Numerical	PRG for ingestion of deer with risk level = 1E-4
C4INGF	Numerical	PRG for ingestion of fish with risk level = 1E-4
C4INGH	Numerical	PRG for ingestion and inhalation with risk level = 1E
C4INGHD	Numerical	PRG for ingestion and inhalation and dermal with ris
C4INGHX	Numerical	PRG for ingestion and inhalation and external with risk level = 1E-4
C4INH	Numerical	PRG for inhalation with risk level = 1E-4

Environmental Restoration Risk Assessment Program, "Risk Assessment Program Data Management Implementation Plan," U.S. Department of Energy Office of Environmental Management, November 1997 https://rais.ornl.gov/documents/tm232.pdf

McMaster University | Library

Social Media Data

- Often contains personally identifying information
- Must be encrypted when stored on an internet connected device
- Should be de-identified + terms and conditions assessed before sharing openly
- Social Media – ethical complexity

More information on the Sherman Centre for Digital Scholarship website:
https://scds.ca/searchable-online-learning/

McMaster University | Library

Publication - Data Sharing and Archiving for qualitative data.

McMaster University | Library

Data Sharing and Archiving – Qualitative Data

A **data repository** is a web platform and storage space for researchers to deposit data for long-term storage and access.

- **Data Sharing**: Open and free data sharing supports research ideals like **transparency, reusability, accessibility, collaboration**, and maximizes the impact and visibility of research.
- **Data Archiving:** What do you plan to do with your data once it's been published? How will you ensure that your data remains accessible (to you and others) long-term?

Photo by NASA on Unsplash.

Photo by Levi Meir Clancy on Unsplash

Why to **not share** qualitative data.

- Framework of established trust.
- Context hard to capture in documentation.
- Data too sensitive or contains PHI
- Concern for data used in bad faith.
- Subverting racialized surveillance or avoiding damage-centered narratives.

McMaster University | Library

Why to **consider sharing** qualitative data.

- Time and resources – for researchers
- Maximizing contributions of research participants
- Alleviate burden on over-researched groups
- Rareness of circumstances

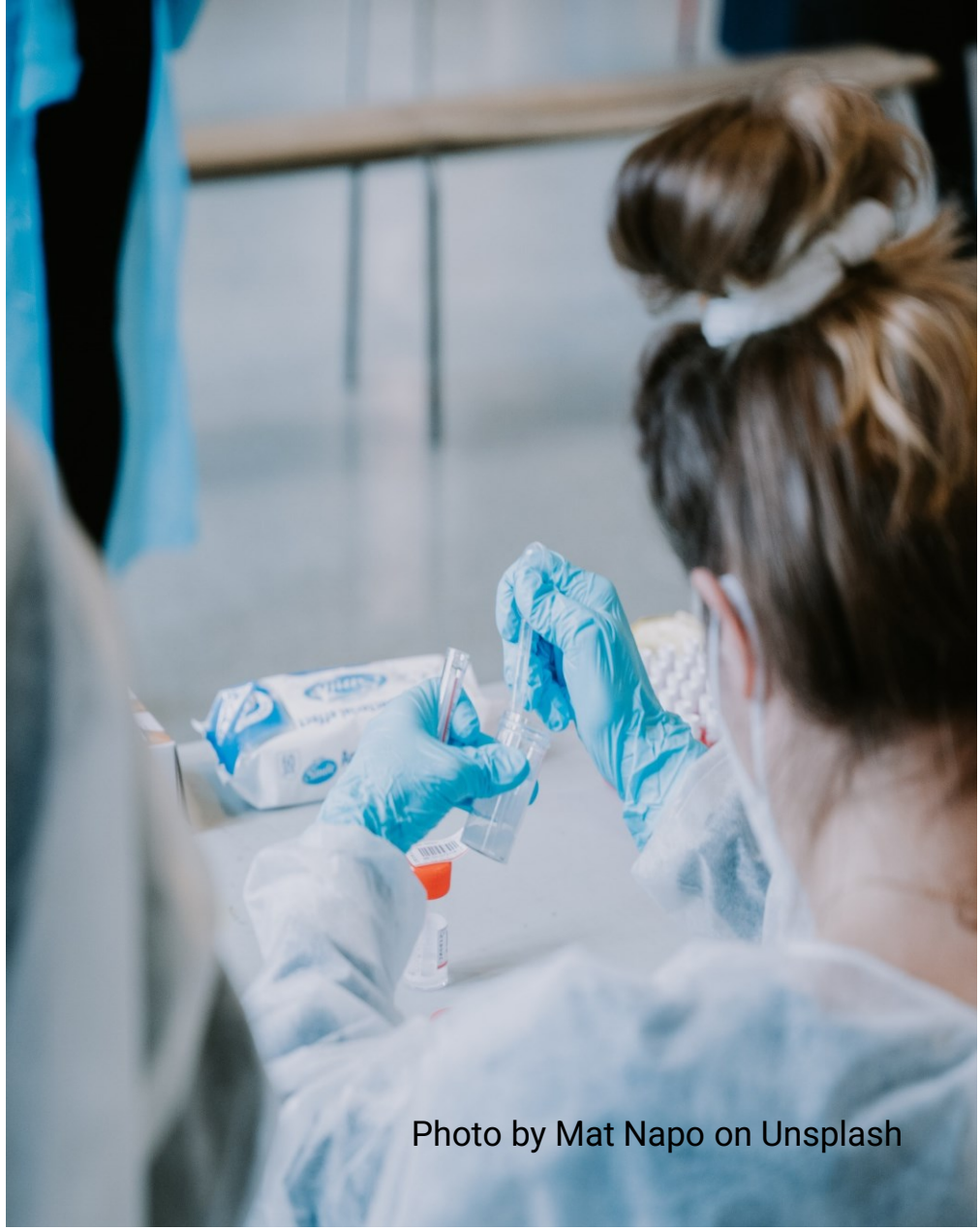


Photo by Mat Napo on Unsplash

Qualitative data **sharing in context**: Queer in Brighton – Dr .Sharon Webb

Won’t asking to share data impact validity?

“Our respondents, both those who did and did not consent to data sharing, indicated that they did not think about data sharing during interviews beyond the initial informed consent process. These findings provide some initial evidence that **data sharing does not reduce the candidness of, or otherwise affect, participants’ responses during qualitative interviews**.”

VandeVusse, Alicia, Jennifer Mueller, and Sebastian Karcher. “Qualitative Data Sharing: Participant Understanding, Motivation, and Consent.” *Qualitative Health Research* 32, no. 1 (January 2022): 182–91. https://doi.org/10.1177/10497323211054058.

Data Repository Decision-Making

Publishing data in a secure data repository is the best way to share and archive data.
Controlled access repositories are best for qualitative data.

Vivli
Non-profit org for sharing clinical data. They enact Data Use Agreement and share data. Budget costs for platform in DMSP.
vivli.org/

Qualitative Data Repository (QDR)
Archive for storing and sharing qualitative data. Carefully curated, tiered restrict access.
qdr.syr.edu/

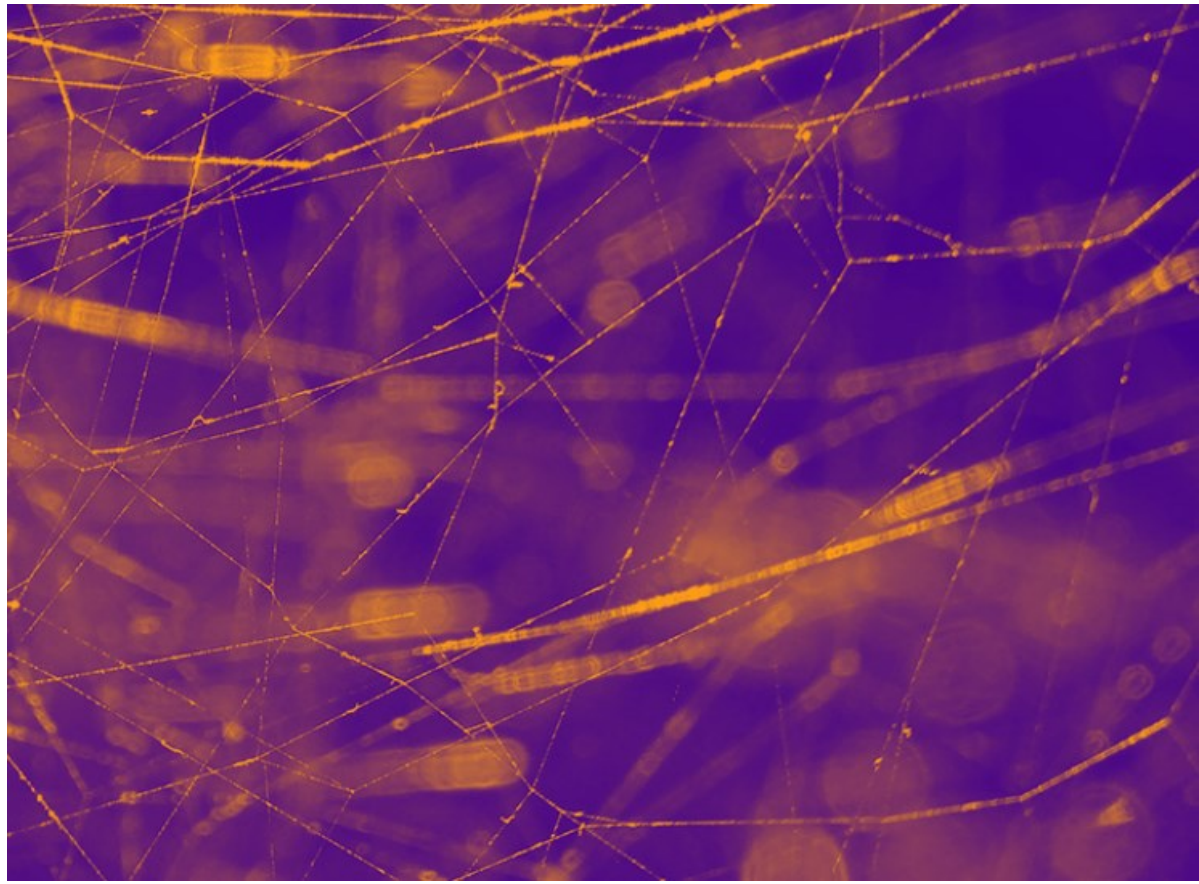
Inter-university Consortium for Political and Social Research (ICPSR)
Training and repository for social science research.
icpsr.umich.edu/

FRDR Sensitive Data Pilot Project
Canada's Federated Research Data Repository (FRDR) has a pilot to add zero-knowledge encryption.
rdm@mcmaster.ca

McMaster University | Library

RDM Community of Practice

- Monthly meetings of people interested in RDM at McMaster
- **Thursday April 27th – 11 AM**
- Arun Keepanasseril - Canadian Hemophilia Assessment and Resource Management System (CHARMS). Arun is also the researcher liaison for REDCap.
- Connect with other researchers practicing RDM across the university!
- https://u.mcmaster.ca/rdm-community



McMaster University | Library

McMaster RDM webpage: rdm.mcmaster.ca

Contact RDM services at: rdm@mcmaster.ca

Upcoming RDM webinars: rdm.mcmaster.ca/events

Recorded RDM webinars: u.mcmaster.ca/learn-rdm

Make an appointment with a Research Data Management Specialist:
u.mcmaster.ca/rdm-appointments

McMaster University Library