ACENET

accelerate discovery

CERC Health Equity & Community Wellbeing Training

January 29-30, 2025

Led by: Meghan Landry, ACENET

A regional partner of the

Digital Research Alliance of Canada

What is research data management?

“Research Data Management (RDM) refers to the storage, security, access and preservation of data produced from a given Scholarly, Research, and Creative (SRC) activity. Data management practices cover the entire lifecycle of the data, from planning the SRC activity to conducting it, and from backing up data as it is created and used to long term preservation of the data after the SRC activity has concluded.”

Taken from Toronto Metropolitan University (TMU)’s Institutional Research Data Management Strategy (2023)

Research Data Management & Tri-Agencies

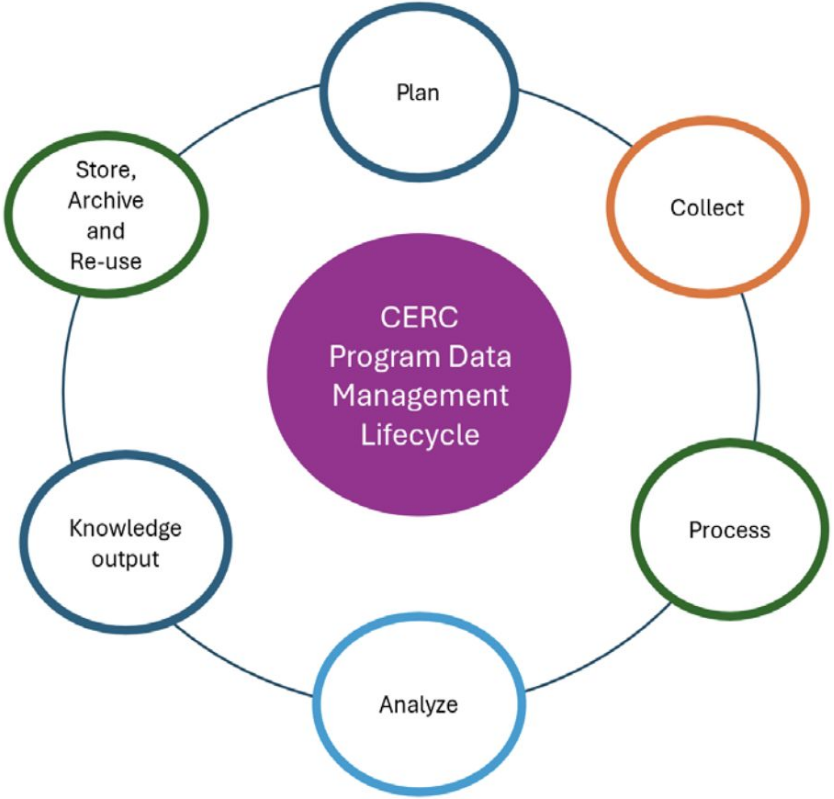
In the spring of 2021, the Canadian Tri-Agencies released the Tri-Agency Research Data Management Policy. This policy outlines three core expectations surrounding RDM for both institutions and researchers:

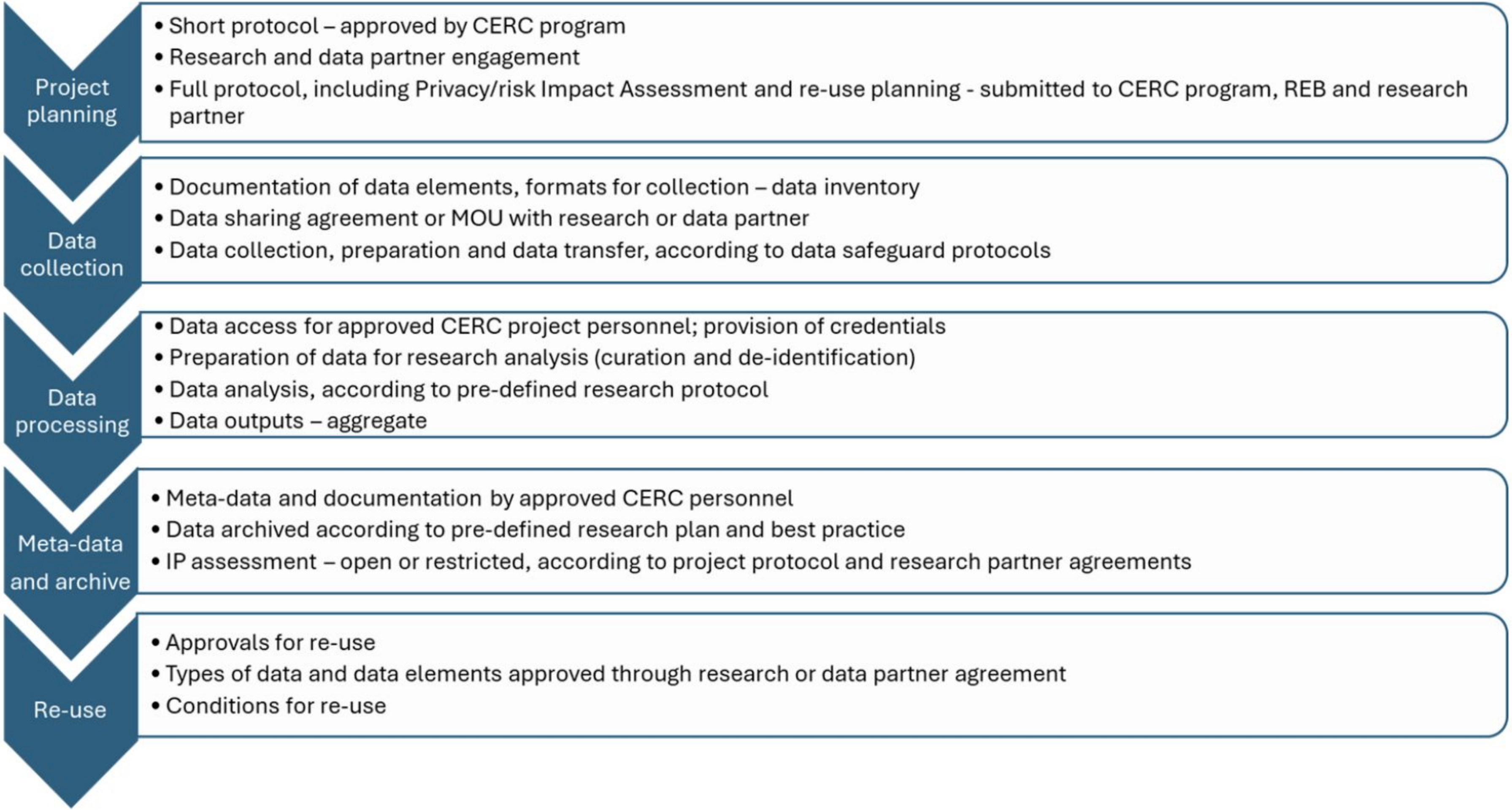
1. institutional strategies
2. data management plans (DMPs)
3. data deposit.

Government of Canada Gouvernement du Canada

The first expectation is to be completed by institutions, while the remaining two – data management plans and data deposit – are the responsibility of researchers.

Research data management lifecycle for HECW


Plan
Collect
Process
Analyze
Knowledge output
Store, Archive and Re-use
CERC Program Data Management Lifecycle



Project planning

- Short protocol – approved by CERC program
- Research and data partner engagement
- Full protocol, including Privacy/risk Impact Assessment and re-use planning - submitted to CERC program, REB and research partner

Data collection

- Documentation of data elements, formats for collection – data inventory
- Data sharing agreement or MOU with research or data partner
- Data collection, preparation and data transfer, according to data safeguard protocols

Data processing

- Data access for approved CERC project personnel; provision of credentials
- Preparation of data for research analysis (curation and de-identification)
- Data analysis, according to pre-defined research protocol
- Data outputs – aggregate

Meta-data and archive

- Meta-data and documentation by approved CERC personnel
- Data archived according to pre-defined research plan and best practice
- IP assessment – open or restricted, according to project protocol and research partner agreements

Re-use

- Approvals for re-use
- Types of data and data elements approved through research or data partner agreement
- Conditions for re-use

The following learning materials are adapted from the University of British Columbia's Introduction to Research Data Management modules

1. Plan

Toronto Metropolitan University

Canada Excellence Research Chair in Health Equity & Community Wellbeing

1. Data partner engagement (Alliance)
2. Responsibilities and Resources
3. Research ethics

Research Ethics

Managing and sharing research data involves various legal, ethical, and intellectual property issues that need to be addressed and respected. These issues may include:

- the protection of personal information,
- the confidentiality of sensitive data,
- the consent of data subjects,
- the ownership of data,
- the attribution of data sources,
- the compliance with relevant laws and regulations, and more.

Source

Responsibilities & Resources

Planning data stewardship, resources and costs helps to avoid confusion, duplication, or conflict among the research team and other stakeholders, to plan ahead and ensure the availability and suitability of the resources for the data management and sharing needs. It also helps to estimate the costs and benefits of the resources and justify their use.

Data stewardship ensures that the data is properly managed, maintained, and used throughout the project lifecycle and beyond. **It is therefore important to define the data stewardship roles and responsibilities of project members during and after the project.**

Source

Responsibilities

These roles and responsibilities should include the following elements:

- The **tasks** that each role will perform
- The **duration** of each role’s involvement in the project
- The **training** that each role will receive to perform their tasks effectively and efficiently
- The **succession plan** that will ensure the continuity and quality of data stewardship in case of staff turnover or role changes among data curators.

Source

2. Collect

1. Documentation of data elements, formats for collection

2. Data collection, preparation and data transfer, according to data safeguard protocols

Toronto Metropolitan University

Canada Excellence Research Chair in Health Equity & Community Wellbeing

Collecting & transforming different data types

Qualitative Data: Typically unstructured or semi-structured, including interviews, open-ended survey responses, focus group discussions, or observational notes.

- Qualitative research focuses on themes, meanings, and patterns. Preparing qualitative data involves organizing and coding it systematically.

Quantitative Data: Structured and numeric, often collected through surveys, experiments, or secondary datasets.

- Quantitative research relies on transforming data for statistical analysis.

Transforming data for research

1. **Data Cleaning (both)**
    a. Remove duplicates, irrelevant information, handle missing values, and correct errors.
    b. Standardize variable names and data formats.
    c. Transcribe audio or video recordings into text for analysis (qual).
    d. Convert qualitative responses into numeric values (quan)

Consider using an open-source tool like OpenRefine, “for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.”

Transforming data for research

2. Data structuring (both)

a. Organize data into rows and columns in a tabular format.
   i. Use software like Excel, SPSS, R, or Python to manage data.

b. Break down large texts into manageable units, such as sentences or paragraphs (qual).
   i. Use software like NVivo or Taguette to import and structure data.

Transforming data for research

3. **Encoding**

a. Assign codes (labels) to specific segments of the data based on themes or concepts.

b. Assign numeric values to categorical variables (e.g., "Male" = 1, "Female" = 2).

c. Scale continuous variables as needed (e.g., normalization or standardization).

d. Organize codes into categories or broader themes.

e. Ensure categories are exhaustive and mutually exclusive, where possible.

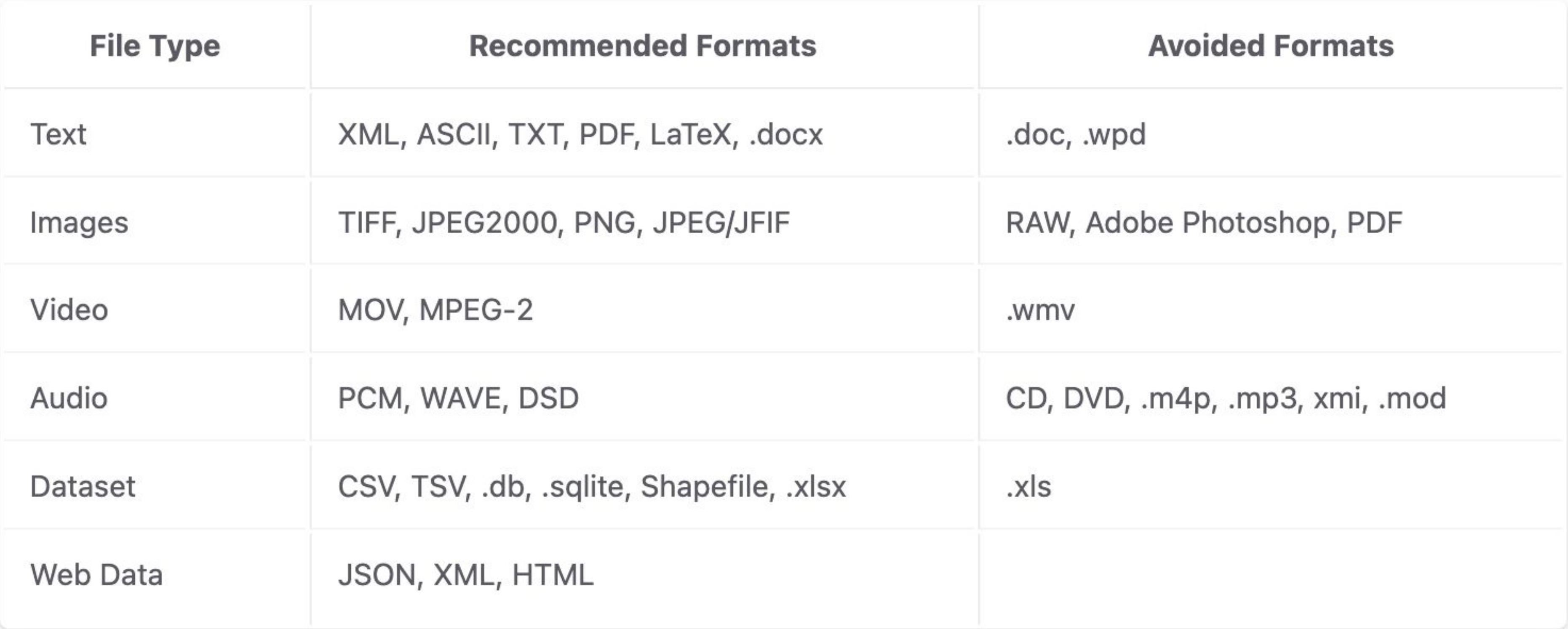
Transforming data for research

4. **Data visualization**

   a. Use tools like word clouds, matrices, or conceptual maps to highlight themes and patterns (qual)

   b. Create charts, histograms, or scatterplots to identify trends and relationships. (quan)

File formats: Proprietary vs. open formats



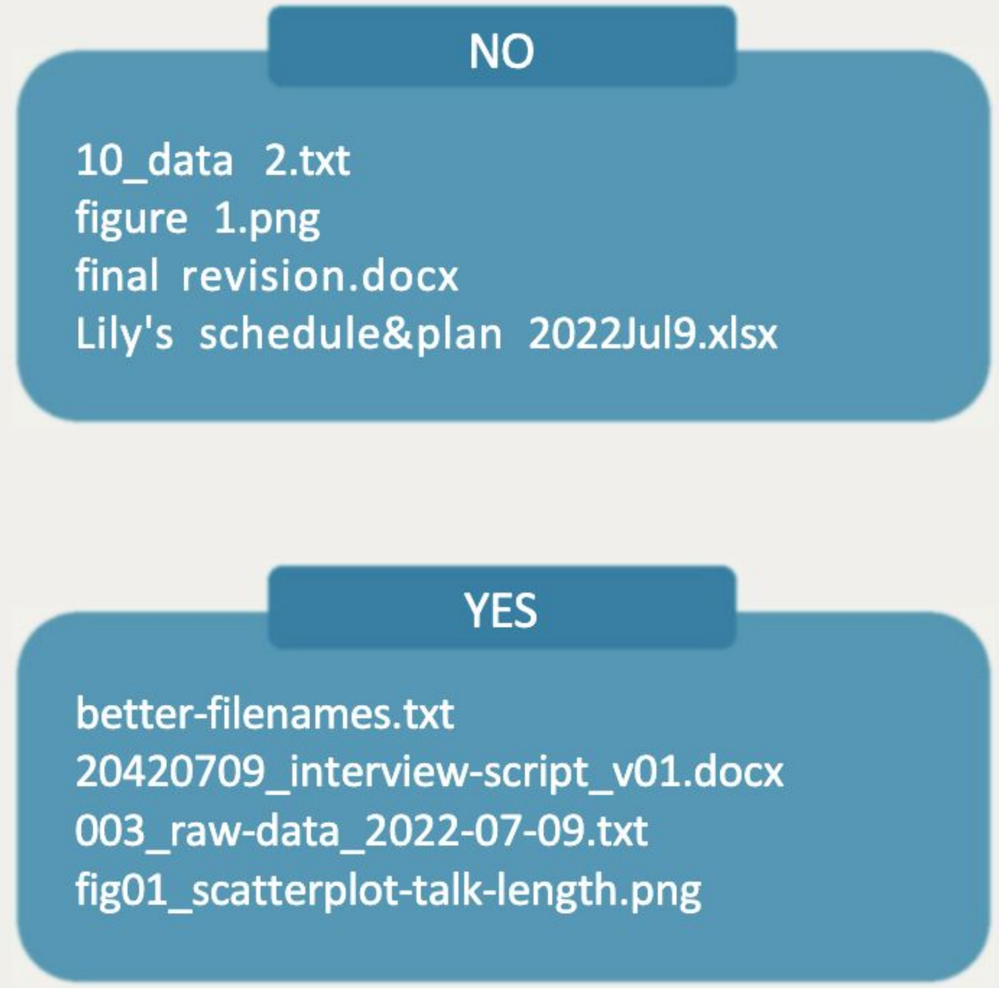
File Type	Recommended Formats	Avoided Formats
Text	XML, ASCII, TXT, PDF, LaTeX, .docx	.doc, .wpd
Images	TIFF, JPEG2000, PNG, JPEG/JFIF	RAW, Adobe Photoshop, PDF
Video	MOV, MPEG-2	.wmv
Audio	PCM, WAVE, DSD	CD, DVD, .m4p, .mp3, xmi, .mod
Dataset	CSV, TSV, .db, .sqlite, Shapefile, .xlsx	.xls
Web Data	JSON, XML, HTML	

Three principles for file naming

1. **Machine-readable:**
   a. Characters in file names are handled correctly by all computer systems
   b. Be consistent with the chosen naming convention
2. **Human readable:**
   a. File names provide concise information.
   b. Names are easily understandable to anyone who accesses them in future (including future you)
3. **Plays well with default ordering:**
   a. Decide at the beginning how you want to sort and search for your files:
      i. Chronological order
      ii. Logical order

Source

- **Alphanumeric characters**
- Use _(**underscore**) or -(**hyphen**)to separate words/numbers (snake case).
- Use **capitalization** to separate words/numbers (camel case).
- Be **consistent**
- Use ISO 8601 standard for dates: **YYYYMMDD or YYYY-MM-DD**
- When using a sequential numbering system, use **leading zeros** to make sure files sort in sequential order. e.g. 001, 002, 010, 011....100,101 ...



NO

10_data 2.txt
figure 1.png
final revision.docx
Lily's schedule&plan 2022Jul9.xlsx

YES

better-filenames.txt
20420709_interview-script_v01.docx
003_raw-data_2022-07-09.txt
fig01_scatterplot-talk-length.png

ACENET
accelerate discovery

Source

Version control

There are many tools for aiding researchers with version control. These tools vary from simpler automated backup systems to platforms with customizable backup and versioning capabilities. If you use any, e.g. Git and Github; The Open Science Framework (OSF) or others - mention and name them in your plan.

Most Alliance federation member organizations regularly offer training in version control.

Source

3. Process

Toronto Metropolitan University

Canada Excellence Research Chair in Health Equity & Community Wellbeing

1. Preparation of data for research analysis (curation and de identification)
2. Data analysis
3. Data outputs

A review of DRAC’s Information Security Roles

Data Custodian: The person who operationally manages one or more sets of data under the direction of the Data Owner or Data Steward.

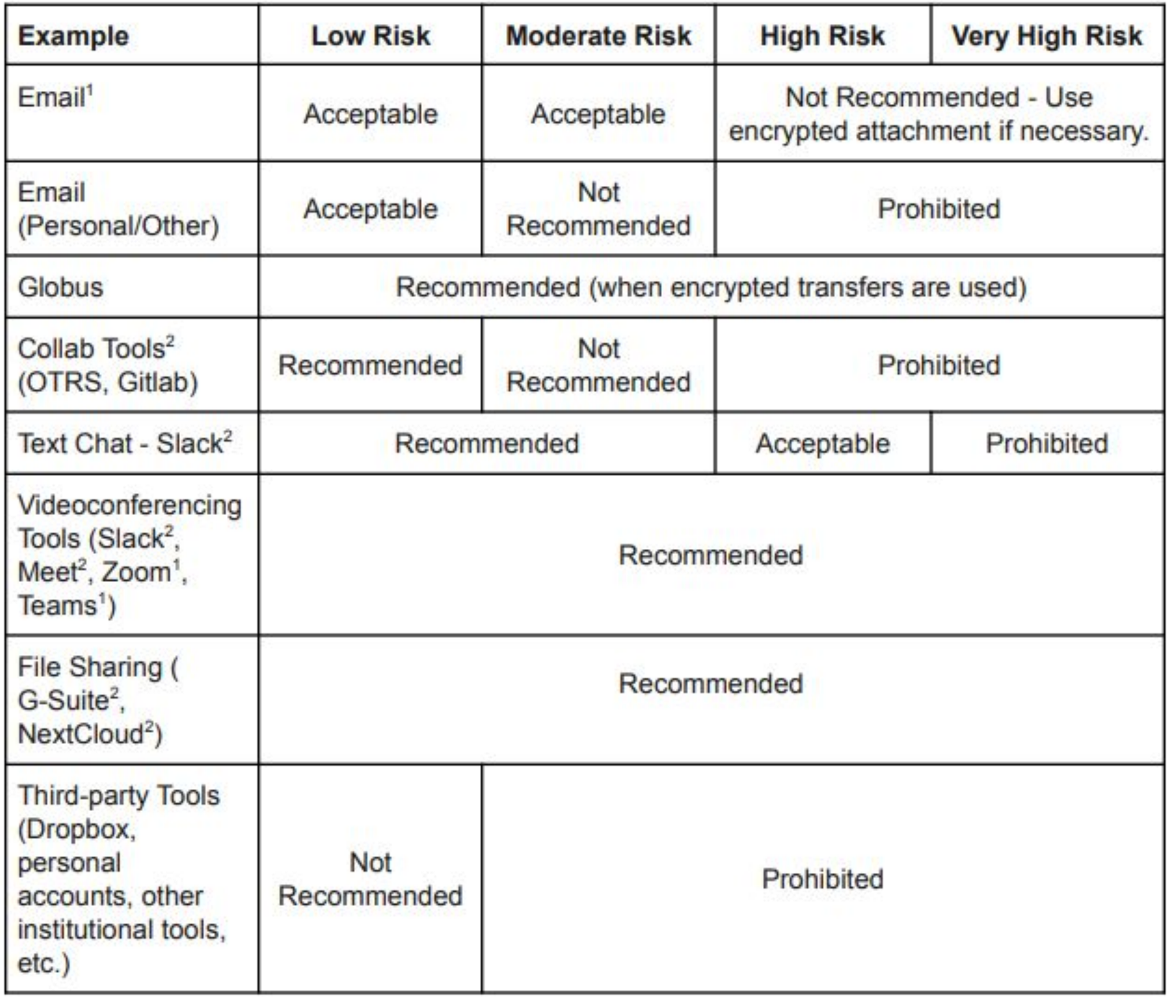
Data Owner: The senior-most role accountable for the data throughout its lifecycle.

Data Steward: The delegate of the Data Owner, a role that oversees the lifecycle of one or more sets of data associated with a National Service. And can make those decisions as delegated by the Data Owner.

Source

A review of the DRAC’s cybersecurity policies

1. **Data Classification Policy**: The purpose of this policy is to establish the methodology for classifying data based on its level of sensitivity, value and criticality to the Alliance Federation.
   a. The **Steward/Owner** is responsible for determining the data security classification of their datasets.
   b. The **Custodian** is responsible for knowing the types of electronic data under their control, the risk their datasets present to the Alliance Federation or its affiliate, its data security classification, and where it is stored.
2. **Data Handling Standard**: This standard details appropriate handling of data based on the data classification defined in the above policy.
   a. Where possible, Moderate, High, or Very High Risk data must be labeled to specify its classification and owner.
   b. The **Data Owner or their assigned Data Steward** must maintain an inventory of all High or Very High Risk data for which they are responsible.



Example	Low Risk	Moderate Risk	High Risk	Very High Risk
Email[1]	Acceptable	Acceptable	Not Recommended - Use encrypted attachment if necessary.	
Email (Personal/Other)	Acceptable	Not Recommended	Prohibited	
Globus	Recommended (when encrypted transfers are used)			
Collab Tools[2] (OTRS, Gitlab)	Recommended	Not Recommended	Prohibited	
Text Chat - Slack[2]	Recommended		Acceptable	Prohibited
Videoconferencing Tools (Slack[2], Meet[2], Zoom[1], Teams[1])	Recommended			
File Sharing (G-Suite[2], NextCloud[2])	Recommended			
Third-party Tools (Dropbox, personal accounts, other institutional tools, etc.)	Not Recommended	Prohibited		

Source
support@ace-net.ca

Working with sensitive data

Human participant data

In Canada, at the federal, provincial, and institutional levels, various legal, policy, and regulatory frameworks govern sensitive data involving humans.

National level: The Tri-Council Policy Statement: [Ethical Conduct for Research Involving Humans (TCPS-2 or the Policy)]

- Joint policy of Canada's three federal research agencies – CIHR, NSERC, SSHRC
- Provides the principles and guidelines that govern the ethical conduct of human participant research in Canadian institutions eligible to receive funding from the three federal agencies noted.

Adapted from Rod, A. B., & Thompson, K. (2023). Sensitive Data: Practical and Theoretical Considerations. *Research Data Management in the Canadian Context: A Guide for Practitioners and Learners.*

However, data can still be used to harm groups or communities.

What is sensitive data?

The Sensitive Data Toolkit for Researchers defines sensitive data as “information that must be safeguarded against unwarranted access or disclosure.”

Examples include:

- Personal information
- Personal health information
- Educational records
- Customer records
- Financial information
- Criminal information
- Geographic information
- Confidential personnel information
- Information that is deemed to be confidential; information entrusted to a person, organization or entity with the intent that it be kept private and access be controlled or restricted.
- Information that is protected by institutional policy from unauthorized access

DRAC’s Sensitive Data Toolkit for Researchers

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

The **Sensitive Data Expert Group** created these tools to help researchers understand how research data are related to the research ethics process, and to advance RDM practices, such as data sharing and deposit, in the context of existing research ethics frameworks.

Part 1: Glossary of Terms for Sensitive Data used for Research Purposes. PDF

Part 2: Human Participant Research Data Risk Matrix. PDF

Part 3: Research Data Management Language for Informed Consent. PDF

How do I know if my data is sensitive?

1. **“Can I Share my Data?” published by DRAC’s COVID-19 Working Group** - a decision tree to alert Canadian researchers to situations where research data derived from human participants either may not be shared publicly or may require some modification before sharing
   a. If YES, what type of data can be shared? Identified, de-identified, or anonymous?

2. **Consult your Research Ethics Board**

Risk & Harm

DRAC's Human Participant Research Data Risk Matrix (Part II of the Sensitive Data Toolkit) - Risk can be determined through consideration of three factors:

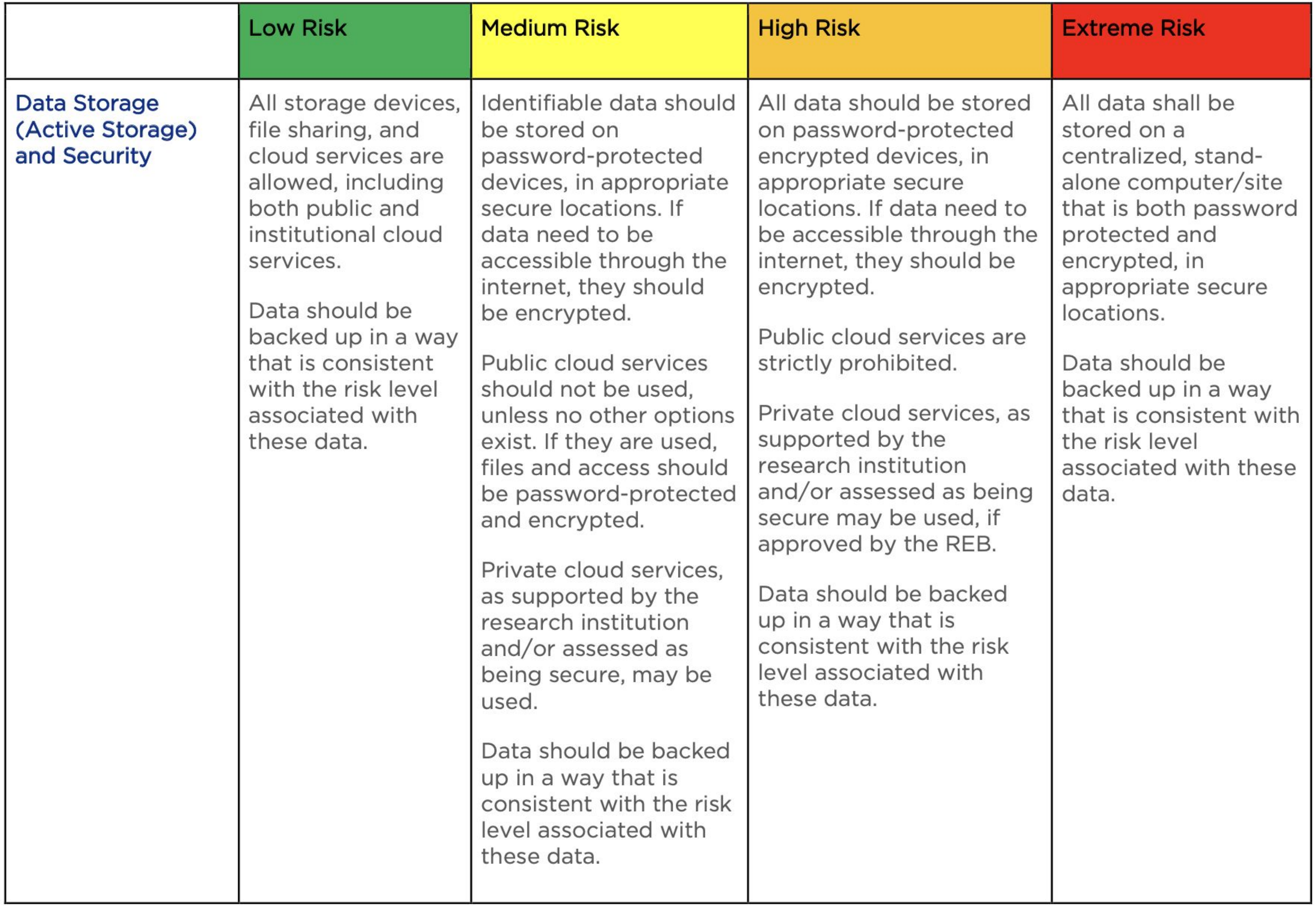
1. identifiability of the data at the time of collection and deposit;

2. vulnerability of the data subjects as individuals or as part of a community or population; and

3. sensitivity of the data in terms of its ability to cause harm - e.g., physical, psychological/emotional, social and legal.

Low Risk
Medium Risk
High Risk
Extreme Risk

Rod, A. B., & Thompson, K. (2023). Sensitive Data: Practical and Theoretical Considerations. *Research Data Management in the Canadian Context: A Guide for Practitioners and Learners.*



	Low Risk	Medium Risk	High Risk	Extreme Risk
Data Storage (Active Storage) and Security	All storage devices, file sharing, and cloud services are allowed, including both public and institutional cloud services. Data should be backed up in a way that is consistent with the risk level associated with these data.	Identifiable data should be stored on password-protected devices, in appropriate secure locations. If data need to be accessible through the internet, they should be encrypted. Public cloud services should not be used, unless no other options exist. If they are used, files and access should be password-protected and encrypted. Private cloud services, as supported by the research institution and/or assessed as being secure, may be used. Data should be backed up in a way that is consistent with the risk level associated with these data.	All data should be stored on password-protected encrypted devices, in appropriate secure locations. If data need to be accessible through the internet, they should be encrypted. Public cloud services are strictly prohibited. Private cloud services, as supported by the research institution and/or assessed as being secure may be used, if approved by the REB. Data should be backed up in a way that is consistent with the risk level associated with these data.	All data shall be stored on a centralized, stand-alone computer/site that is both password protected and encrypted, in appropriate secure locations. Data should be backed up in a way that is consistent with the risk level associated with these data.

Refer to Human Participant Research Data Risk Matrix when assessing data storage options

support@ace-net.ca

Protecting active sensitive data

Taken from TMU's How to Protect Confidential Data:

1. Work from a shared drive or server
    a. As an alternative to Google Drive, CCS also provides Central File and Print Services (CFAPS), which are personal and/or shared-access folders on file servers, free of charge for TMU faculty and staff.
2. Use two-factor authentication
3. Password protect your devices
4. Seek permission to transport data
    a. Consider secure tools to transfer data, like Globus
5. Choose secure devices
6. Encrypt your folders

Data encryption - how it works

- Encryption distorts or scrambles data so it can only be read by authorized people.
- The only way to unscramble data is by entering a unique decryption key.
  - On some devices such as computers running Microsoft Windows 10, the unique encryption key is integrated with your device login process, which unscrambles your data automatically.
  - Other devices such as select USB drives, require users to enter a unique key or password before unscrambling the data.
- On any device, it is recommended that you keep a copy of your decryption key in a safe place to unscramble your data whenever required by your system.

Source

Enabling encryption

There are a number of encryption softwares available, including:

- BitLocker for Windows PC
- FileVault for Mac
- VeraCrypt for both PC and Mac

De-identification (or anonymization):

“The act of changing individual-level data to decrease the probability of disclosing an individual’s identity. This can involve masking direct identifiers (e.g., name, phone number, address) as well as transforming (e.g., recoding, combining) or suppressing indirect identifiers that could be used alone or in combination to-identify an individual (e.g., birth dates, geographic details, dates of key events). If done correctly, de-identification minimizes and therefore mitigates risk of re-identification of any data shared or released.”

Taken from DRAC’s Glossary of Terms for Sensitive Data used for Research Purposes

De-identification Guidance

Taken from DRAC's De-identification Guidance report

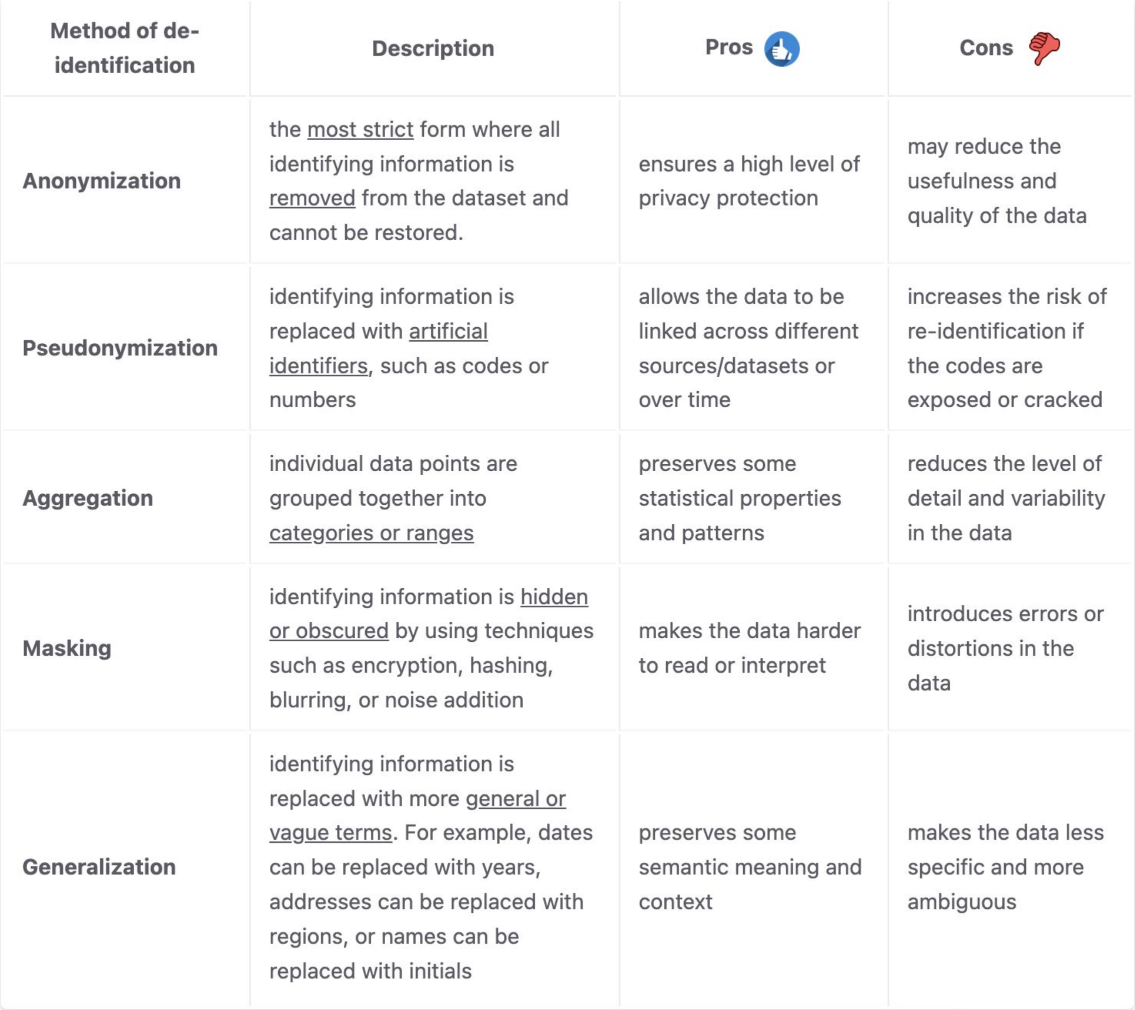
1. Identify and remove direct identifiers
   a. You may either record this personal information in a separate document, spreadsheet, or database and link this to the other data points via a series of codes that can be removed before publishing or choose to delete the identifying data points entirely at the end of the project.

2. Identify and Evaluate Indirect or Quasi-Identifiers based on Perceived Risk and Utility
   a. Quasi-identifiers may not be identifying on their own but can be disclosive in **combination.**
   b. Observe possible combinations
   c. Assess these combinations using K-anonymity

Considerations for de-identifying qualitative data

General advice:

1. Avoid asking for identifying information in the first place.
2. Make de-identification a part of the process of informed consent
3. Use pseudonyms and change identifying details to protect anonymity.
4. If necessary, remove blocks of sensitive text or edit out portions within audio-visual data.



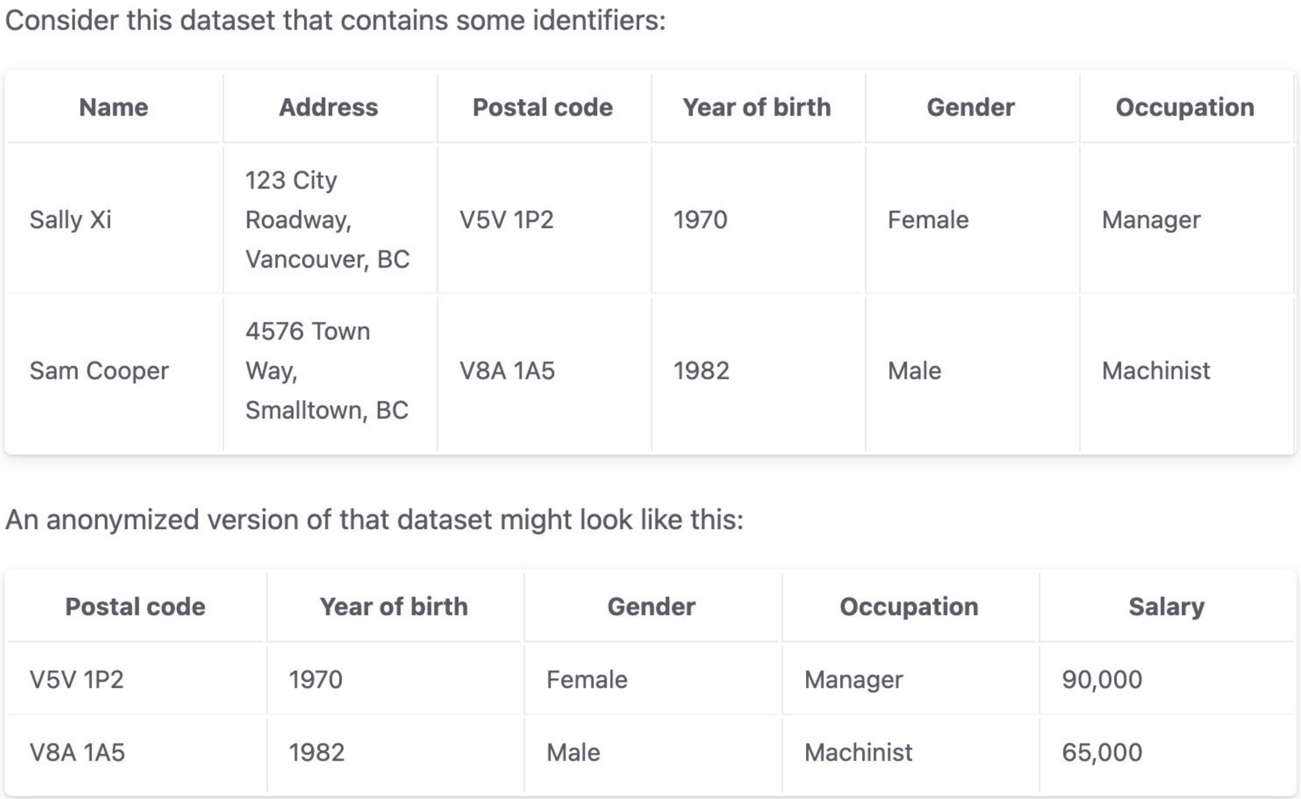
Method of de-identification	Description	Pros 👍	Cons 👎
Anonymization	the most strict form where all identifying information is removed from the dataset and cannot be restored.	ensures a high level of privacy protection	may reduce the usefulness and quality of the data
Pseudonymization	identifying information is replaced with artificial identifiers, such as codes or numbers	allows the data to be linked across different sources/datasets or over time	increases the risk of re-identification if the codes are exposed or cracked
Aggregation	individual data points are grouped together into categories or ranges	preserves some statistical properties and patterns	reduces the level of detail and variability in the data
Masking	identifying information is hidden or obscured by using techniques such as encryption, hashing, blurring, or noise addition	makes the data harder to read or interpret	introduces errors or distortions in the data
Generalization	identifying information is replaced with more general or vague terms. For example, dates can be replaced with years, addresses can be replaced with regions, or names can be replaced with initials	preserves some semantic meaning and context	makes the data less specific and more ambiguous

ACENET
accelerate discovery

Source

support@ace-net.ca

Example of anonymization



Consider this dataset that contains some identifiers:

Name	Address	Postal code	Year of birth	Gender	Occupation
Sally Xi	123 City Roadway, Vancouver, BC	V5V 1P2	1970	Female	Manager
Sam Cooper	4576 Town Way, Smalltown, BC	V8A 1A5	1982	Male	Machinist

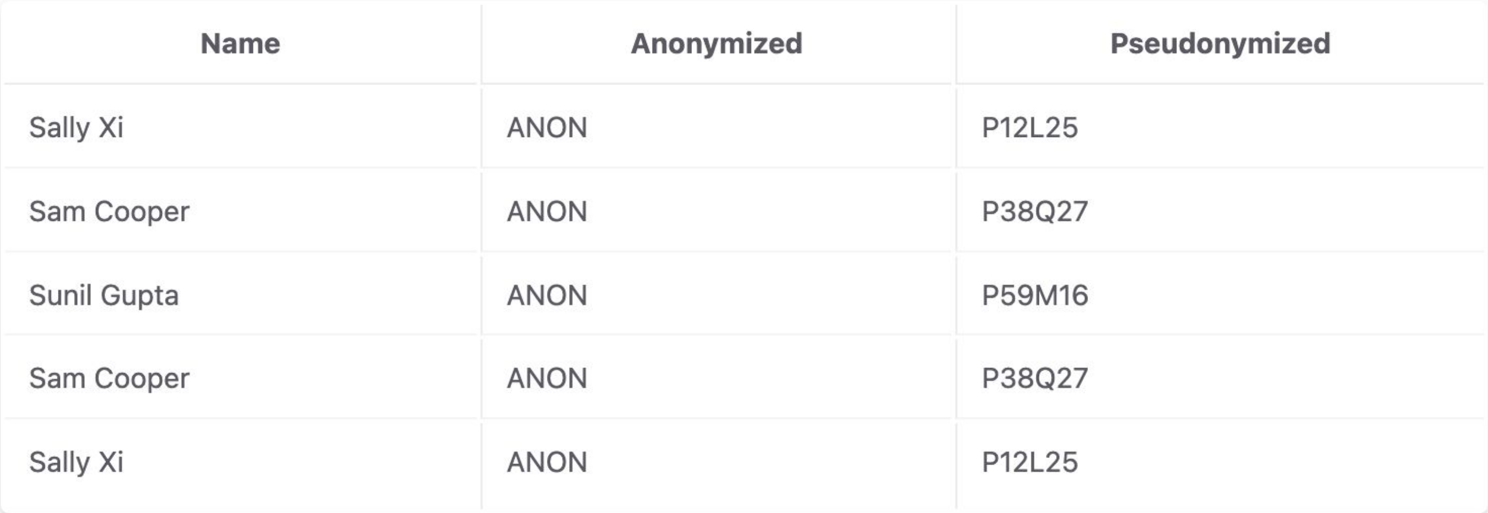
An anonymized version of that dataset might look like this:

Postal code	Year of birth	Gender	Occupation	Salary
V5V 1P2	1970	Female	Manager	90,000
V8A 1A5	1982	Male	Machinist	65,000

Source

Example of pseudonymization

Data pseudonymization can preserve the linkability and utility of the data. Linkability means that the data can be connected to the same individual or entity across different datasets or over time.



Name	Anonymized	Pseudonymized
Sally Xi	ANON	P12L25
Sam Cooper	ANON	P38Q27
Sunil Gupta	ANON	P59M16
Sam Cooper	ANON	P38Q27
Sally Xi	ANON	P12L25

Source

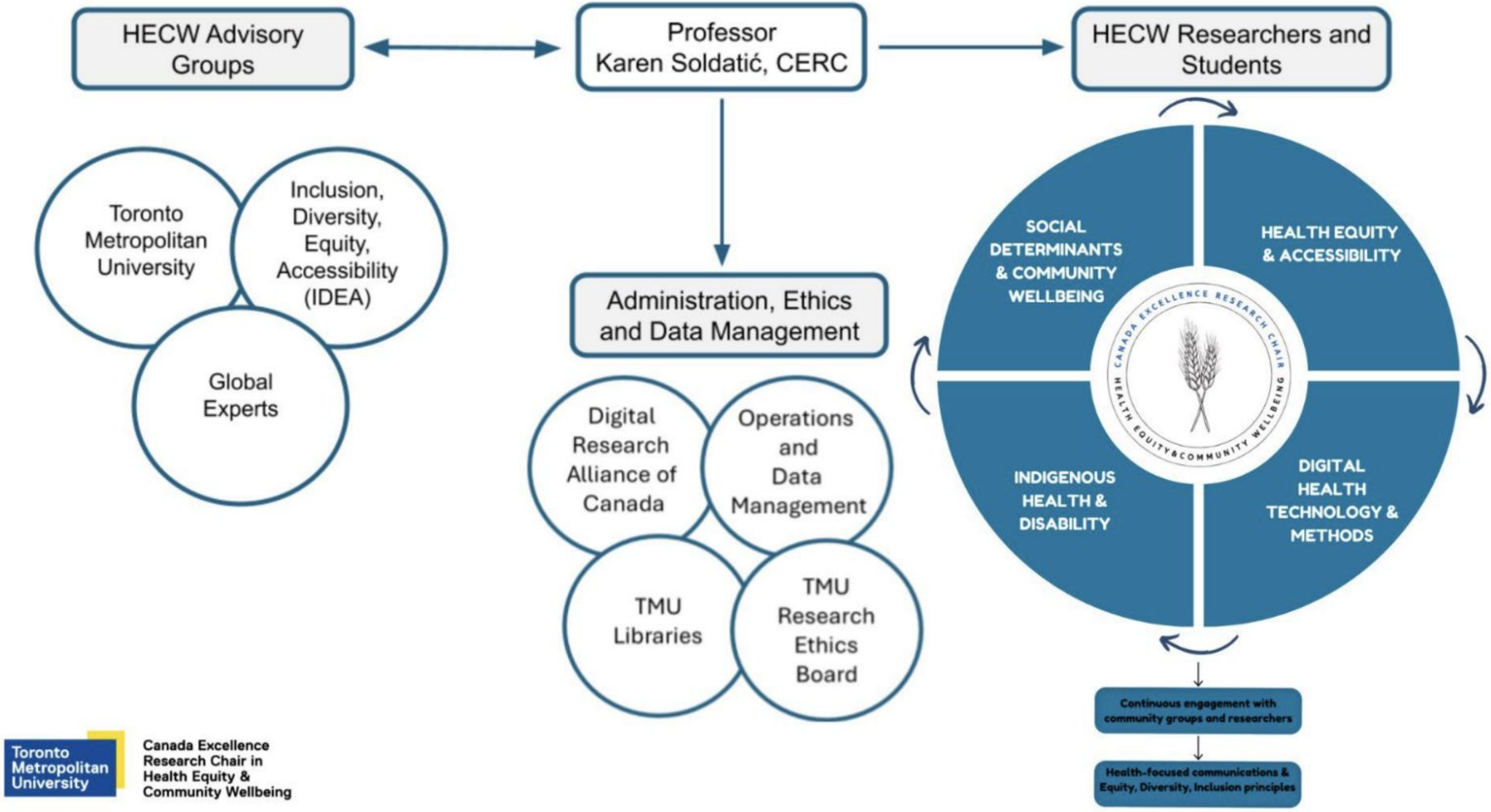
De-identification tools

Researchers are increasingly using algorithm-based tools to help anonymize their data and manage the risk of re-identifying their anonymized data. An example of an anonymization tool would be:

ARX open source data anonymization software

Source

Figure 1: Overarching CERC Health Equity and Community Wellbeing Governance Structure



HECW Advisory Groups

Professor Karen Soldatić, CERC

HECW Researchers and Students

Toronto Metropolitan University

Inclusion, Diversity, Equity, Accessibility (IDEA)

Global Experts

Administration, Ethics and Data Management

Digital Research Alliance of Canada

Operations and Data Management

TMU Libraries

TMU Research Ethics Board

SOCIAL DETERMINANTS & COMMUNITY WELLBEING

HEALTH EQUITY & ACCESSIBILITY

INDIGENOUS HEALTH & DISABILITY

DIGITAL HEALTH TECHNOLOGY & METHODS

CANADA EXCELLENCE RESEARCH CHAIR HEALTH EQUITY & COMMUNITY WELLBEING

Continuous engagement with community groups and researchers

Health-focused communications & Equity, Diversity, Inclusion principles

Toronto Metropolitan University

Canada Excellence Research Chair in Health Equity & Community Wellbeing

Data Stewardship Roles & Responsibilities based on Governance Structure

DRAC’s Data Handling Standards

Create a spreadsheet based on the governance structure with:

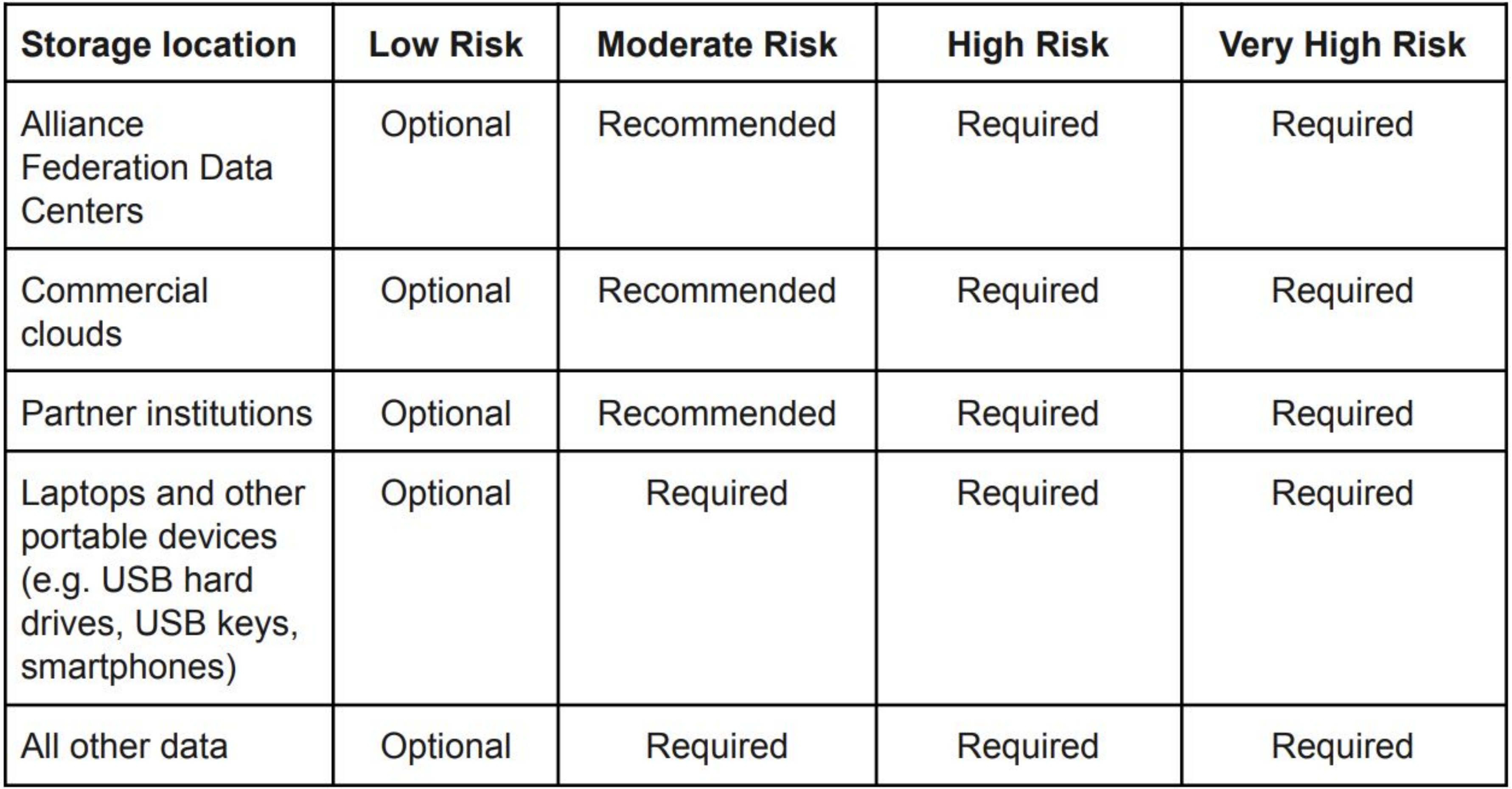
- **Roles in the program and corresponding roles in data stewardship** (user, custodian, steward, owner, etc.)
- **Data Classification** (Low risk, medium risk, high risk, extreme high risk)
- **Location/System**
- **Medium/Format** (Digital document, Database, Hardcopy)
- **A General Description**

Protocols for securely transferring data from one location to another (3rd party data)

Government of Canada - Guidance on securely configuring network protocols (ITSP.40.062)

Data transmission involves a copy of data being moved from one place to another. When data is transferred from one security realm/domain to another it must be secured in accordance with its data classification.

- All data should be transferred in accordance with industry best practices for encryption in transit
- It is recommended that all data be transferred through an encrypted channel whenever possible.



Storage location	Low Risk	Moderate Risk	High Risk	Very High Risk
Alliance Federation Data Centers	Optional	Recommended	Required	Required
Commercial clouds	Optional	Recommended	Required	Required
Partner institutions	Optional	Recommended	Required	Required
Laptops and other portable devices (e.g. USB hard drives, USB keys, smartphones)	Optional	Required	Required	Required
All other data	Optional	Required	Required	Required

Source

4. Metadata & archive

1. Metadata and documentation
2. Preparing data for archiving

Toronto Metropolitan University

Canada Excellence Research Chair in Health Equity & Community Wellbeing

ACENET
accelerate discovery

What is metadata?

Metadata is often described as “data about data” and helps answer the questions of who, what, when, where, why. This descriptive data is essential for creating FAIR and open data, and ensuring that the datasets you preserve will be accessible for many years to come.

Metadata makes it easier for researchers to:

- share their data,
- publicize their data,
- locate and retrieve data sets from others.

Metadata best practices

1. **Descriptive**: Descriptive metadata describes the content and context of your data at both the dataset and item level.

   Examples: title, author, keywords

2. **Administrative**: Administrative metadata includes information needed to use the data.

   Examples: software requirements, copyright, licensing

3. **Structural**: Structural metadata describes how different datasets relate to one another, or any processing or formatting steps that were undertaken.

   Examples: Information about the relationship between datasets in a database, file formats

Source

Creating a README file

A README is a guide to your dataset and is usually a plain text file to **maximize its usability and long-term preservation potential**.

The purpose of a README is to assist other researchers to **understand your dataset, its contents, provenance, licensing and how interact with it**.

README files are generally named **README, readme.txt or read-me.md** and are included as component of a dataset.

Core elements of any README include:

- Contact information for the researcher
- The use license for your data
- The context of your data collection
- Your data collection methods (protocols, sampling, instruments, coverage, etc.)
- The structure of files
- Naming conventions for files, if applicable
- Your sources used
- Your quality assurance work
- Any data manipulations or modifications
- Data confidentiality and permissions
- The names of labels and variables
- Explanations of codes and classifications

Codebooks & Data Dictionaries - important for Qual Software

McGill’s Codebook Cookbook: A guide to writing a good codebook for data analysis projects in medicine

- **Variable name**: The name or number assigned to each variable in the data collection. For survey data, try to name variables after the question numbers - e.g., Q1, Q2b, etc. [In above example, H40-SF12-2]
- **Variable label**: A brief description to identify the variable for the user. ["SF12 - ASSESSMENT OF R'S GENERAL HEALTH"]
- **Question text**: Where applicable, the exact wording from survey questions. ["In general, would you say your health is . . ."]
- **Values**: The actual coded values in the data for this variable. [1, 2, 3, 4, 5]

Source

Codebooks continued

- **Value labels**: The textual descriptions of the codes. [Excellent, Very Good, Good, Fair, Poor]
- **Summary statistics**: Where appropriate and depending on the type of variable, provide unweighted summary statistics for quick reference.
- **Missing data**: Where applicable, the values and labels of missing data. Missing data can bias an analysis and is important to convey in study documentation. Remember to describe all missing codes, including "system missing" and blank. [e.g., Refusal (-1)]
- **Universe skip patterns**: Where applicable, information about the population to which the variable refers, as well as the preceding and following variables. [e.g., Default Next Question: H00035.00]

README Files, Codebooks, & Data Dictionaries

README files and Codebooks/Data Dictionaries are critical for transparency and reproducibility because they allow others to easily understand the contents of your directory and data without needing to ask the creator. This is especially helpful when working with a group or sharing directories with others.

Two types of files needed to store all metadata:

- **README file** which resides in our root directory and elaborates on the contents of our folder structure, discusses how, where, and who did the data collection.
- **DATA-DICTIONARY file** that resides in our data directory and elaborates on how our data variables are defined and described.

Source

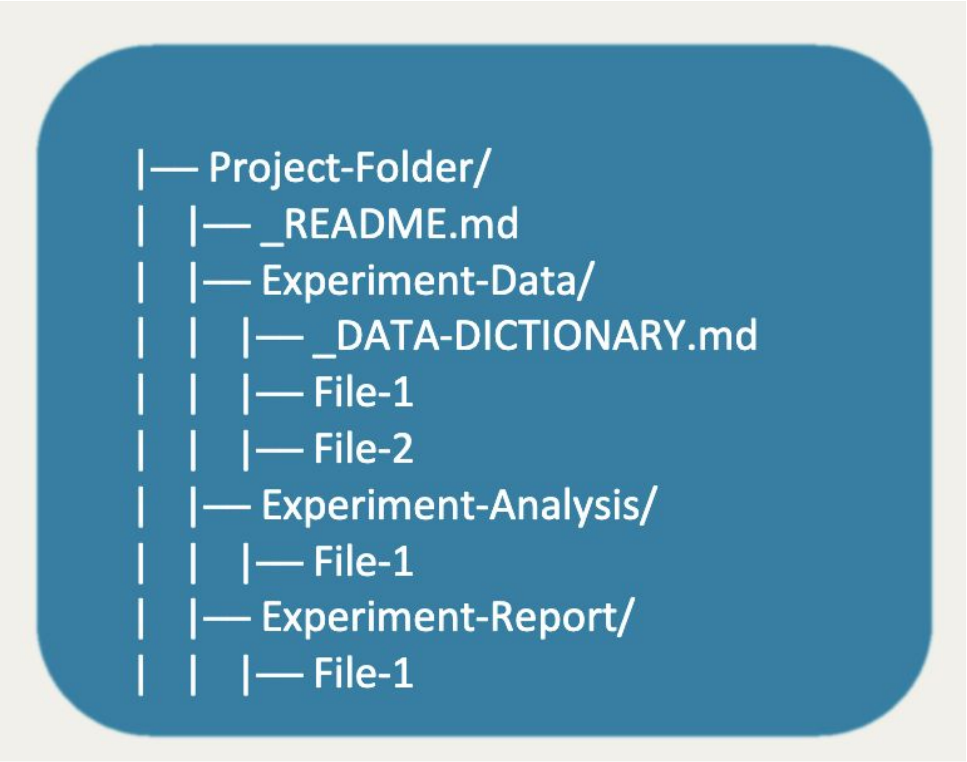
File Directory Hierarchies

A typical directory structure is composed of:

- **a root directory** (i.e. top-level folder)
- **subdirectories** (i.e. subfolders), and
- **relevant files**.

Directory names are frequently followed by a slash/to differentiate them from files.

Usually, we separate data, analysis, and reports into stand-alone subdirectories under the project’s root directory. The structure may look like this:



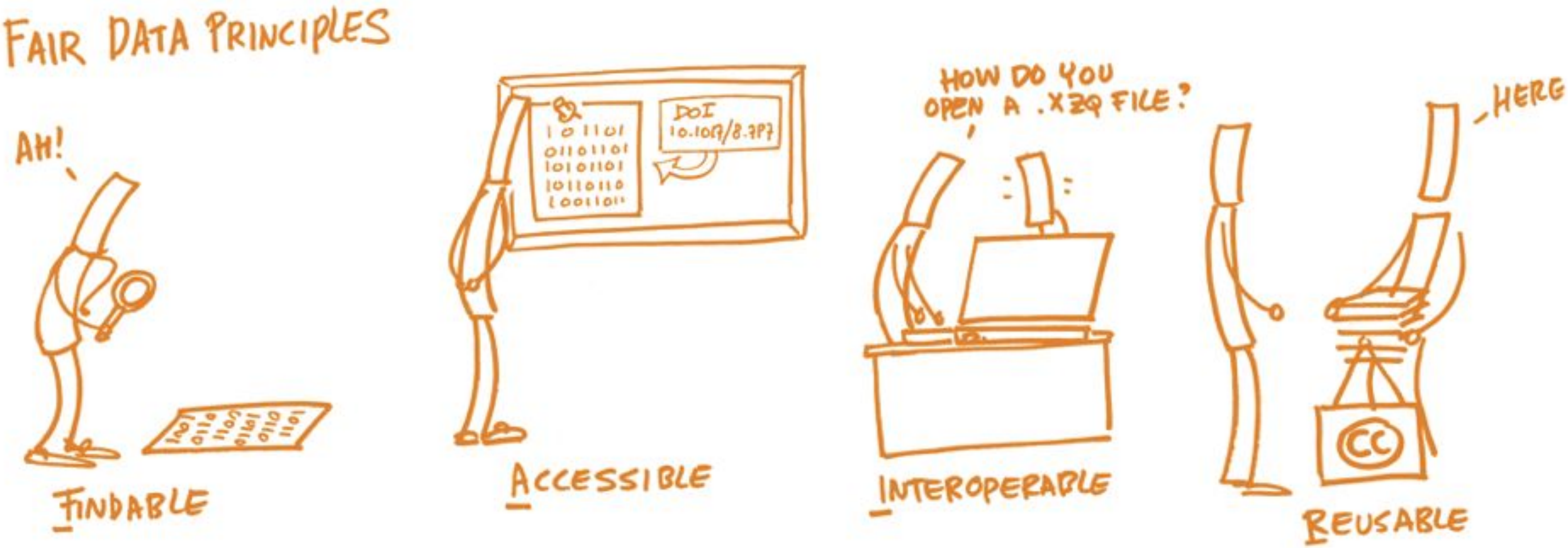
```
|— Project-Folder/
|   |— _README.md
|   |— Experiment-Data/
|   |   |— _DATA-DICTIONARY.md
|   |   |— File-1
|   |   |— File-2
|   |— Experiment-Analysis/
|   |   |— File-1
|   |— Experiment-Report/
|   |   |— File-1
```

5. Share, Archive, & Reuse

Toronto Metropolitan University

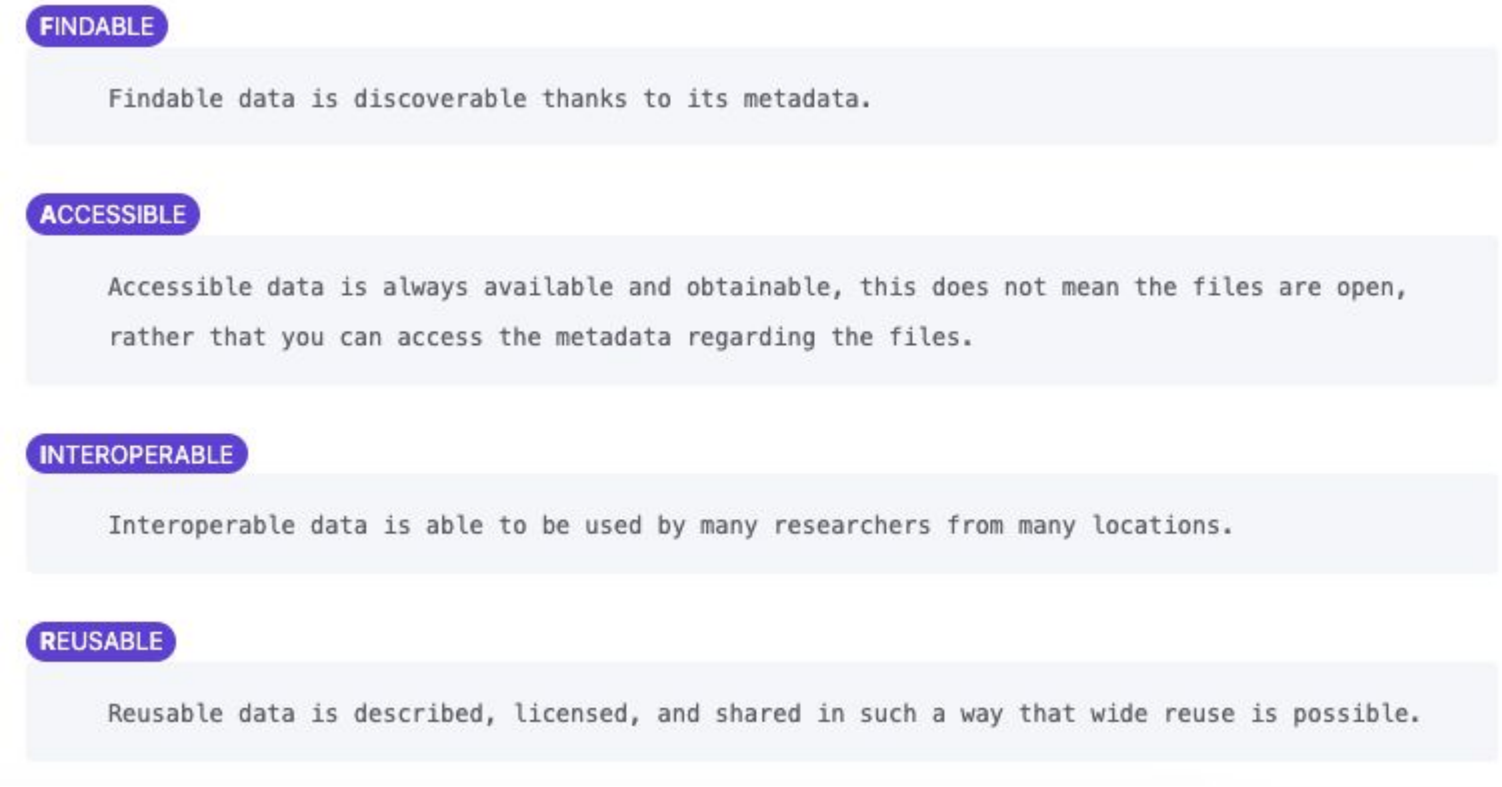
Canada Excellence Research Chair in Health Equity & Community Wellbeing

1. A review of FAIR Data Principles and CARE Principles for Indigenous Data Governance
2. Trustworthy repositories
3. Data destruction



ACENET
accelerate discovery

Source
support@ace-net.ca



FINDABLE

Findable data is discoverable thanks to its metadata.

ACCESSIBLE

Accessible data is always available and obtainable, this does not mean the files are open, rather that you can access the metadata regarding the files.

INTEROPERABLE

Interoperable data is able to be used by many researchers from many locations.

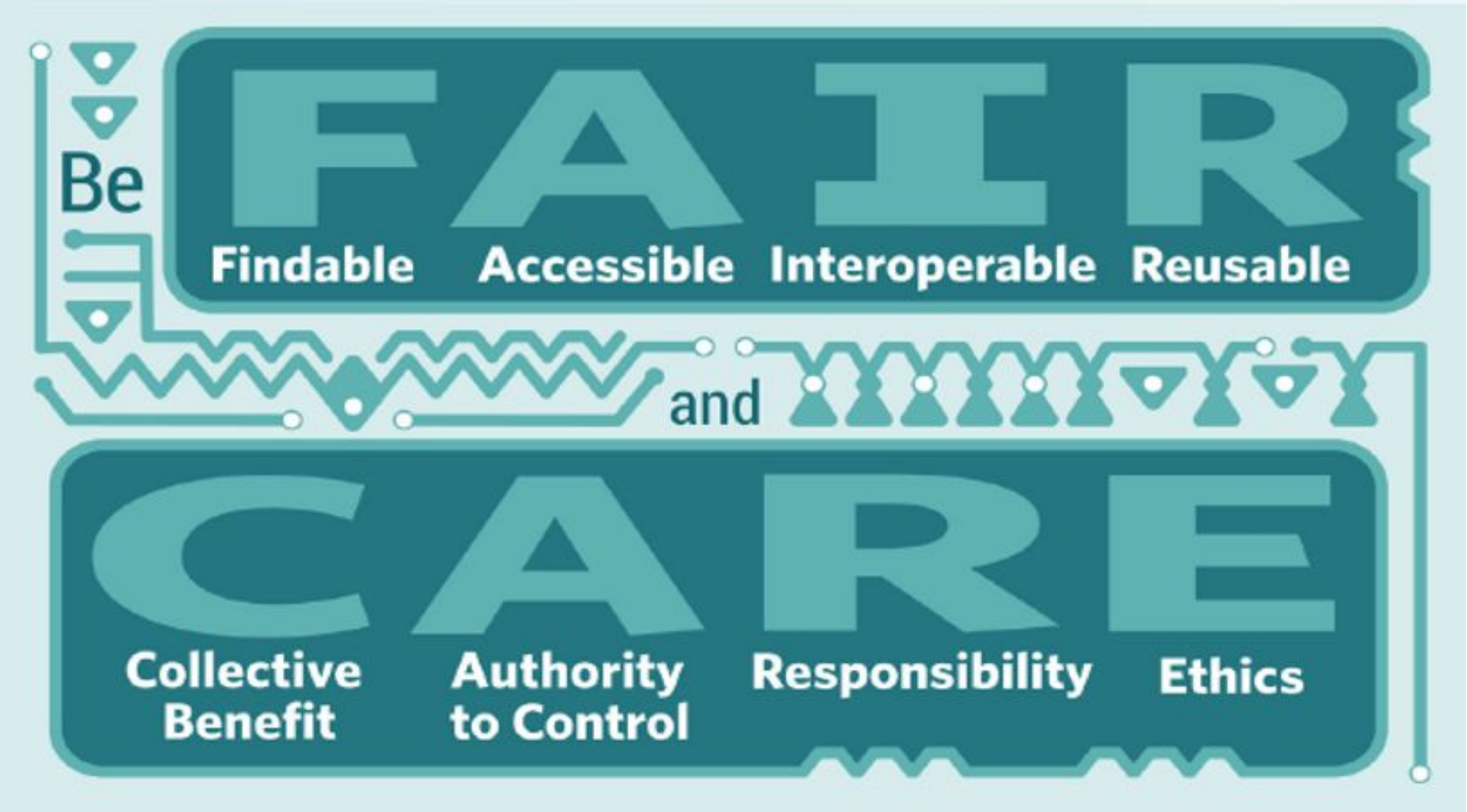
REUSABLE

Reusable data is described, licensed, and shared in such a way that wide reuse is possible.

Source

Note on FAIR data

All data can be FAIR, but not all FAIR data is open. OpenAIRE states that data should be “as open as possible, as closed as necessary.” Not all data can be fully open, but it should still be findable at the metadata level.


Be
FAIR
Findable
Accessible
Interoperable
Reusable
and
CARE
Collective Benefit
Authority to Control
Responsibility
Ethics

CARE Principles for Indigenous Data Governance

Collective Benefit – "Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data."

Authority to Control – Indigenous people have the right and authority to control their data.

Responsibility – Researchers working with Indigenous Peoples have a responsibility to support Indigenous Peoples rights.

Ethics – "Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem."

Data preservation

The preservation process encompasses:

- converting data into sustainable formats,
- storing it in secure repositories,
- viruses control,
- providing comprehensive metadata and documentation.

When considering what data should be preserved, it’s important to note that not all data you create needs to be preserved. In some areas of science, it is often less expensive to resequence the data sample than to store all the data, especially in the preservation layer. Factors such as the value of your data and funding requirements should be taken into consideration.

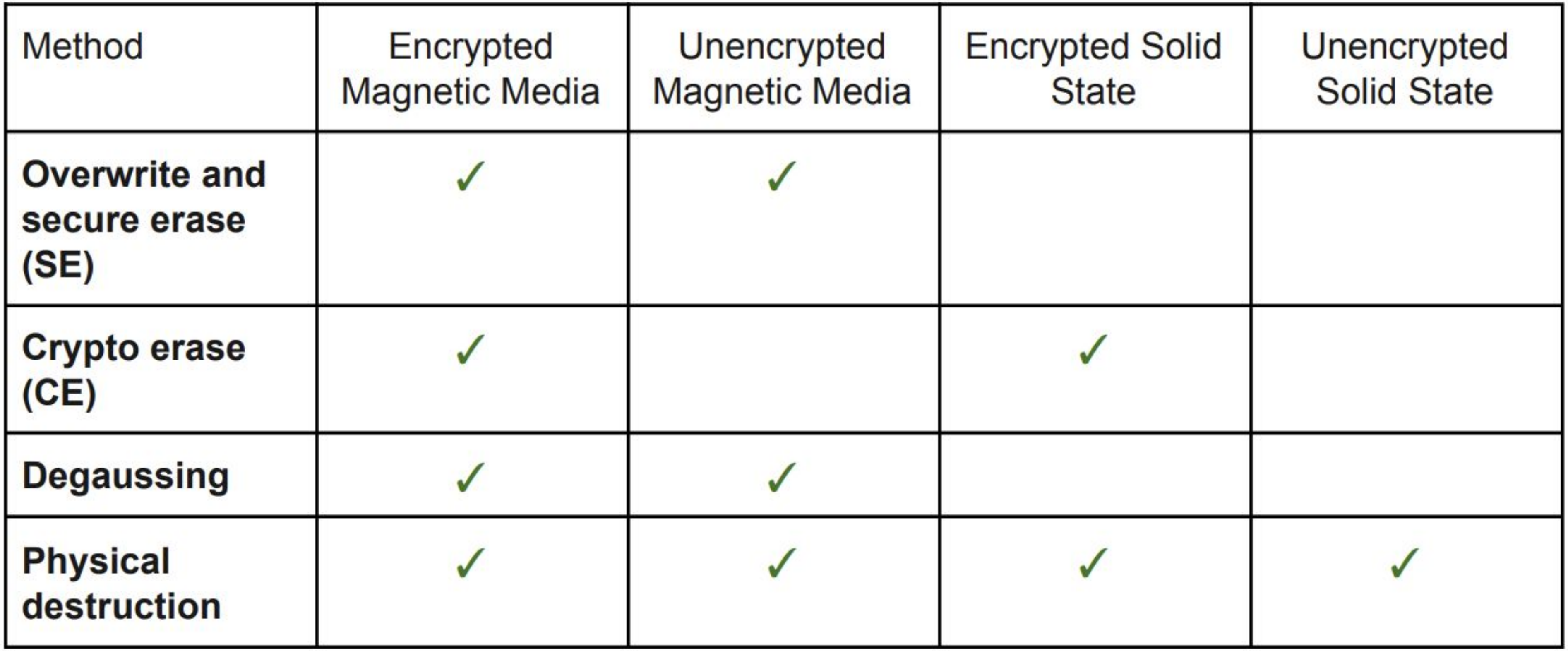
Selecting a data repository from TMU Libraries

Placing your data into a repository allows it to be saved after the life of a research project and makes sharing easier. There are a variety of repositories suited for different needs. Your repository choice may be based on data requirements, discipline as well as journal and funder requirements.

1) **Toronto Metropolitan University Dataverse**
2) **Search for a repository on re3data**: re3data.org is a registry of research data repositories. This resource may help you to identify a repository for your data. You can browse by data type and subject.
3) **PLOS Data Repository Recommendation Guide**: Academic journals may require authors to deposit data related to an article. PLOS ONE has a data sharing requirement.

Data Disposal of High/Extreme Risk Data

- Should be outlined in data sharing agreements and consent forms
- All data must be retained as long as required by applicable regulation and/or policy. Once data is no longer required it must be destroyed, including cases involving the reuse of storage devices.
- Refer to https://cyber.gc.ca/en/guidance/sanitization-and-disposal-electronic-devices-itsap40006 for more information
  - (note that "Erase and factory reset" is insufficient for data destruction).



Method	Encrypted Magnetic Media	Unencrypted Magnetic Media	Encrypted Solid State	Unencrypted Solid State
Overwrite and secure erase (SE)	✓	✓		
Crypto erase (CE)	✓		✓	
Degaussing	✓	✓		
Physical destruction	✓	✓	✓	✓

Source

Preserving sensitive data in Canada: no data repositories in Canada currently allow for the deposit of raw or identifiable sensitive data

LibreQDA and Taguette

LibreQDA is a software modelled off Taguette - After uploading documents, users can highlight words, sentences, or paragraphs and tag them with the codes you create. All the work you do in Taguette is completely exportable, including tagged documents, codebooks, highlights for a specific tag, highlights for all tags, and a list of tags with their descriptions.

A spin on the phrase "tag it!", Taguette is a free and open source qualitative research tool.

After uploading documents, users can highlight words, sentences, or paragraphs and tag them with the codes you create.

More about LibreQDA

LibreQDA is a software modelled off Taguette - and positioned as an open source competitor for NVivo (which is proprietary). It is currently in its late phase of BETA testing, developed by a small team at the Université de Sherbrooke.

Eventually, Taguette is going to stop being updated by the sole dev on the project. So it’s great that we have LibreQDA in the pipeline! And it’s Canadian!

LibreQDA is hosted on a server managed by Calcul Quebec, a partner organization of the Digital Research Alliance of Canada

LibreQDA

Both software are free, open-source qualitative research tools that can be used to accomplish specific steps in the analysis of interviews or other corpuses:

- Importing a variety of text files (.pdf, .docx, .odt, .txt and more)
- Coding words, sentences or paragraphs using codes manually set by the user
- Finding and analyzing themes of interest
- Exporting a subset of coded selections or the project as a whole

Follow-up and going forward

How can you continue to engage our services, expertise, and infrastructure?

- **One-on-one virtual consultations**
- **Creating a NextCloud** account: https://docs.alliancecan.ca/wiki/Nextcloud
- **Submitting tickets** to our technical desk: https://alliancecan.ca/en/services/advanced-research-computing/technical-support/getting-help
- **Documentation**: https://docs.alliancecan.ca/wiki/Technical_documentation
- **Training opportunities**: https://alliancecan.ca/en/services/dri-training