DATA MANAGEMENT PLANS

January 28, 2021

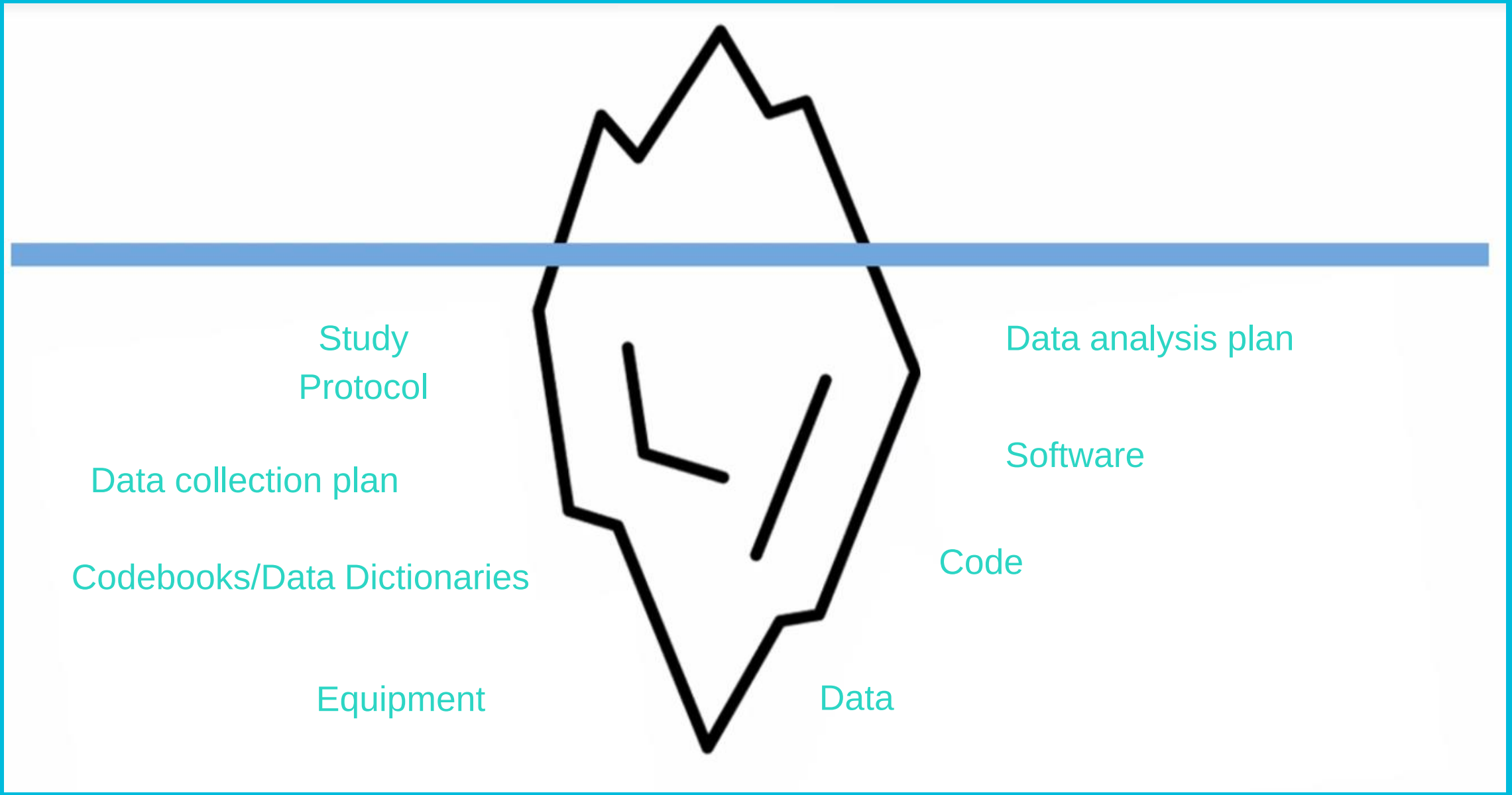
Research Services & Digital Strategies Unit, Libraries

University of Manitoba

Publication

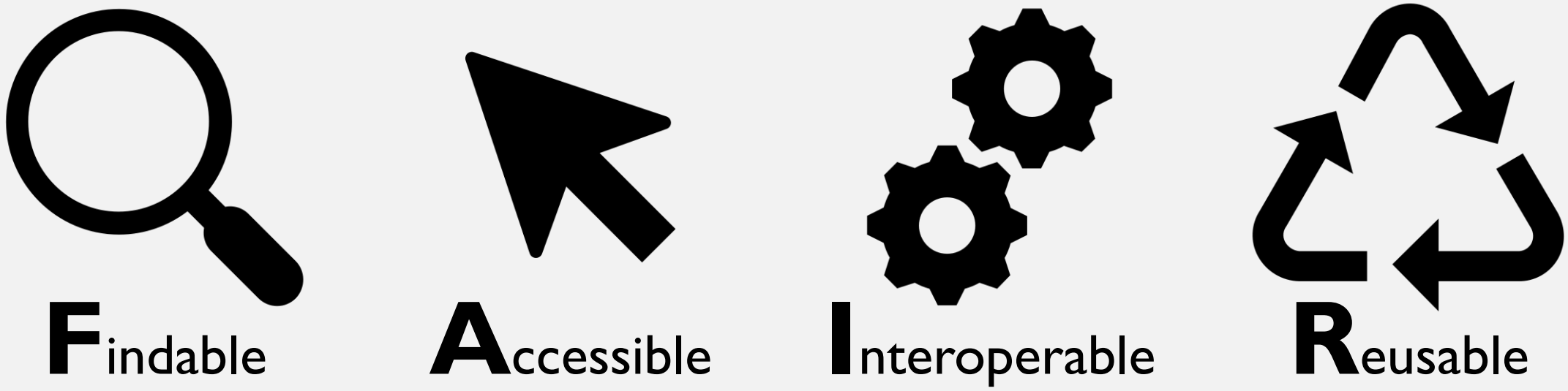
Courtesy: Kevin Read, U. Saskatchewan

Publication



Courtesy: Kevin Read, U. Saskatchewan

FAIR PRINCIPLES



A set of principles to ensure that data are shared in a way that enables and enhances reuse by humans and machines

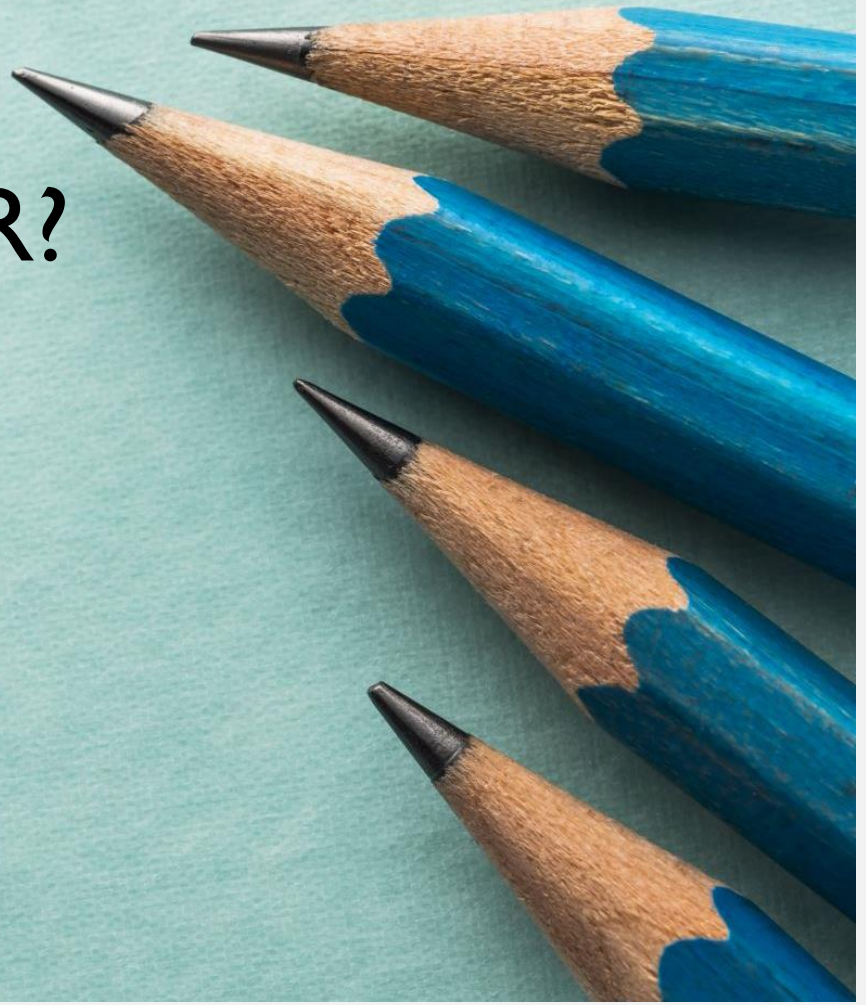
GUIDING PRINCIPLES FOR HUMAN RESEARCH

Source: Global Indigenous Data Alliance

Source: First Nations Information Governance Centre

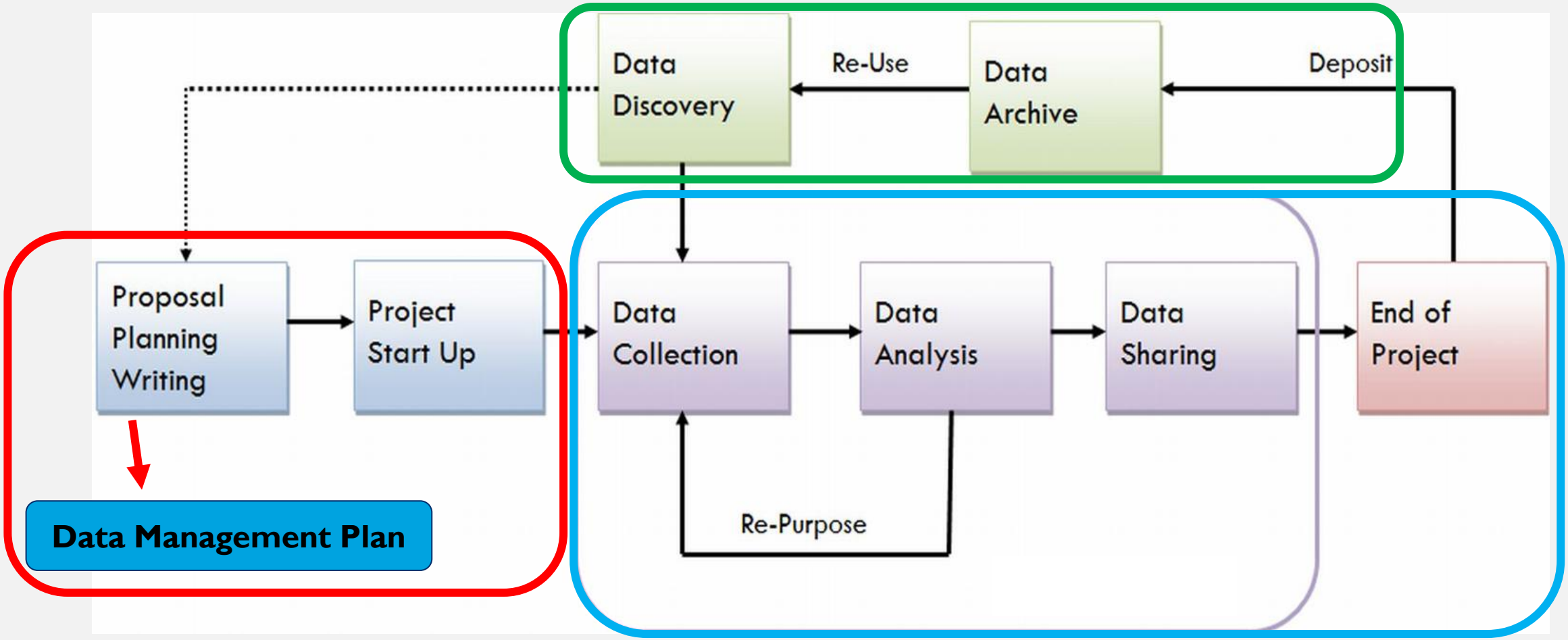
How do I make my data FAIR?

PLANNING



***RDM** = managing data throughout all phases of the research lifecycle…*

Through planning, active research, and beyond



Data Life Cycle

Source: "*E-Science and the Life Cycle of Research*", Humphrey, C. (2006)

portage

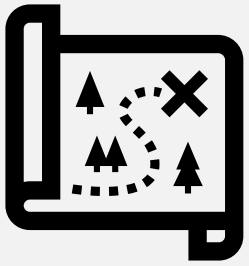
What is a Data Management Plan (DMP)?

- Describes **what** data you expect to acquire or generate during the course of a research project and **why**
- Explains **how** you will manage, describe, analyze, and store your data and **who** will be responsible
- Details **when** and **where** you will deposit your data and **how** it will be shared

- **DMPs are living documents, meant to be reviewed and updated**

https://communitylivingstmarys.ca/services/community-development-and-planning-services/

WHY A DMP?

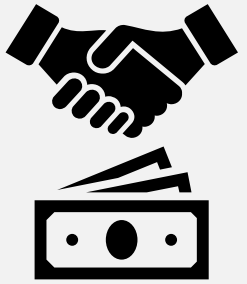


Map out future DM needs

Communicate with Team

Identify questions, areas for improvement

Record of Intentions



Funder Mandate

DMP MANDATES (OR EQUIVALENT)

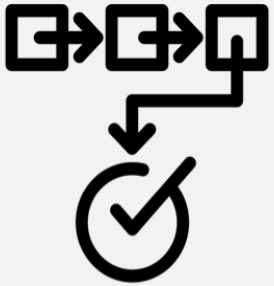
CURRENT

FUTURE?

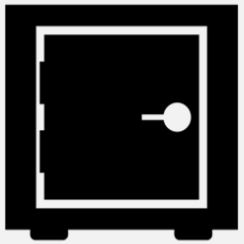
Funder Policies

CIHR IRSC | NSERC CRSNG | SSHRC CRSH

Institutional Strategy



Data Management Plans



Deposit

DRAFT Tri-Agency Research Data Management Policy For Consultation

Institutional Strategy

RDM Strategy Template

Data Management Plans

DMP Assistant

National, online, bilingual, Data Management Planning Tool

New version imminent
New discipline-specific Exemplars & Templates

Deposit

National, Multi-disciplinary Repository Options

\+ University of Manitoba

UM Dataverse
MSpace

portage

Portage DMP Assistant 2.0

Tool

Templates

- Data Collection
- Documentation & Metadata
- Storage & Backup
- Preservation
- Sharing & Reuse
- Responsibilities & Resources
- Ethics & Legal Compliance

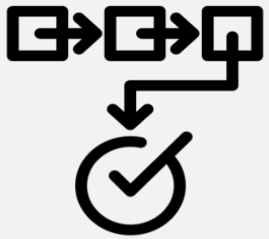
~ Seven Sections

+ Guidance

UNIVERSITY OF ALBERTA

portage

- **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



DISCIPLINE-BASED DMP EXEMPLARS

SUMMARY OF THE DMP QUESTIONS

Review this list and reflect on your own research – have you told yourself your own research story?

University of Manitoba

Components of a Data Management Plan

- Data Collection -
  - Define data types: textual, tabular, graphical, numerical
  - List the file formats that will be used/created: proprietary, open standard. Software or tools required to read and/or view the data
  - Naming conventions for versions, data architecture: documentation in READ ME file, root name
- Documentation and Metadata -
  - Define the documentation accompanying the data generated
  - Outline best practices to ensure accuracy, consistency of documentation
  - Define Metadata standard(s) to be used
- Storage and Backup (during project) -
  - Define backup process - see best practice for backup
  - Outline total storage requirements (e.g. megabytes, gigabytes, terabytes) and length of retention time
  - Access and location: details of access (protection of sensitive data) and documentation
- Data Preservation (after project) -
  - Using deposit with a permanent identifier assigned is preservation best practice
  - Elaborate on retention: duration, format, location, access parameters
  - Describe preparation processes: data selection, de-identification, data formats
- Data Access and Reuse -
  - Stipulate the version of data (raw, processed, analyzed, final)
  - Method of storage and access - resource and access considerations
  - Levels of access: free to read v. ability to reuse
- Responsibilities and Resources -
  - Describe the various types of access, duration that will be granted, and how it will be documented
  - Outline the costs related to the data management lifecycle, including equipment, software, dissemination (including any journal article processing charges) - see the Data Management Costing Tool for estimating what those costs may be
- Ethics and Legal Compliance -
  - Sensitive data: describe your management processes to ensure security and restrict access
  - Strategies for reuse: your data selection criteria, documentation; method(s) to enable access e.g. deposit
  - Identify any/all ethical, legal and intellectual property considerations that may override access, and/or reuse to the data and what processes you have undertaken to manage them

QUICK START TO DMP ASSISTANT

1. Researcher creates a (free) account.
2. Selects a template, usually from the institution to which they belong.
3. Start working through the template: progress and versions are saved; any and/or plans are retained in the account.
   - Collaborators can work on the same DMP
   - English and French is available
4. When plan is complete, can be exported: pdf, csv, html, txt, docx, xml, json

https://libguides.lib.umanitoba.ca/RDMGuide/plan

https://padlet.com/szwajcera/UM_DMPfaqs

RELATED PRESENTATIONS

Data Deposit
https://youtu.be/KI3KqXbzRAA

Data Sharing and Creative Commons Licenses https://youtu.be/HgludPgJolA

Data Management Plans: A How-To
https://youtu.be/fzuaYT8ofWM

Navigating Open Publishing
https://youtu.be/E3LBPtldmqo

Yes, researchers who want to become a little more self-sufficient in handling their data will have to learn some things. But no researcher has to know or be everything. Every conversation between a researcher and a data librarian will make the both of them a little wiser: the librarian learns more about the research that is going on at their institution, the researcher learns what to take into account and why. And all it costs is a little time.

Mariëtte van Selm, Information Specialist (University Library), University of Amsterdam – State of Open Data 2020

University of Manitoba

THANK YOU

Andrea Szwajcer

andrea.szwajcer@umanitoba.ca

Research Services Librarian

RISK MATRIX - LOW

From: **Sensitive data Toolkit for Researchers-** Part 2: Human Participant Research Data Risk Matrix
https://doi.org/10.5281/zenodo.4060159

Low

Data that could be public, does not reasonably identify individuals, groups; does not contain confidential, private, sensitive data; subjects not defined as vulnerable

- **Consent** – notification that data will be made available for future use
- **Collection** – found in public sources; researchers do not know identities of participants; methods are indirect (e.g. surveys, observational); no direct or indirect identifiers collected
- **Analysis** – no restrictions; adhere to REB-approved protocol (and documentation/script contained within)
- **Storage** – all devices, sharing, cloud-services allowed; backup consistent for this level
- **Mobility/Sharing** – shared via email and all cloud services
- **Deposit and Access** – unrestricted within reasonable timeframe; secondary data use does not require REB approval
- **Data Retention & Destruction** – may be retained indefinitely for discovery, access, archival

RISK MATRIX - MEDIUM

From: **Sensitive data Toolkit for Researchers-** Part 2: Human Participant Research Data Risk Matrix
https://doi.org/10.5281/zenodo.4060159

Medium

Identifiers collected have been stripped; may contain originally collected data as confidential, private, sensitive data; subjects not defined as vulnerable

- **Consent –** notification that data will be made available for future use; opt-out for deposit should be considered
- **Collection –** researcher may know identities of participants; may have promised confidentiality through informed consent; methods of collection wide-ranging, involve direct interaction with participants ***Majority of human research will fall into this category***
- **Analysis –** direct identifiers replace asap with linking code; consent forms should be stored separately from research data; only research team members should have access to identifiable data
- **Storage –** identifiable data stored on password-protected devices, in appropriate, secure locations; if accessible via internet, needs to be encrypted; private cloud services, as supported by the research institution and/or assessed as secure, may be used; backup should be consistent with risk level
- **Mobility/Sharing –** encrypted and password-protected files can be shared via email and institution-approved cloud services or collaboration sites
- **Deposit and Access –** data from subject who opt out should be separated from data to be deposited; de-identified data should be deposited unrestricted within reasonable timeframe; secondary data use requires REB approval
- **Data Retention & Destruction** – may be retained indefinitely for discovery, access, archival

RISK MATRIX - HIGH

From: **Sensitive data Toolkit for Researchers-** Part 2: Human Participant Research Data Risk Matrix
https://doi.org/10.5281/zenodo.4060159

High

Identifiers remain and/or (re) identification possible/probable; contains confidential, private, sensitive data; subject may be vulnerable in context of research and may be harmed in case of breach

- **Consent** – notification that data will be made available for future use; request for permission to share and/or deposit data clearly indicated in consent form or process; provide options for areas of future research
- **Collection** – researcher may know identities of participants; may have promised confidentiality through informed consent; methods of collection wide-ranging, involve direct interaction with participants; direct identifiers may or may not be collected, but indirect identifiers collected sufficient to render participants identifiable (triangulation of data potential)
- **Analysis** – direct identifiers replace asap with linking code; consent forms should be stored separately from research data; indirect identifiers should be coded, if possible; data should not be analyzed in public space where data could be observed on a device or by other means.
- **Storage** – identifiable data stored on password-protected, encrypted devices, in appropriate, secure locations; if accessible via internet, needs to be encrypted; public cloud services strictly prohibited; private cloud services, as supported by the research institution and/or assessed as secure, if approved by REB; backup should be consistent with risk level
- **Mobility/Sharing** – restricted data shall only be shared with research team members as specified in the approved protocol; files shall be encrypted, and password protected.
- **Deposit and Access** – data from subject who opt out should be separated from data to be deposited; de-identified data should be deposited with restricted access to be evaluated by data custodian; data may be separated into sets depending on use cases that participants have agreed to via informed consent (e.g. use for this study, studies in this subject area or any use); secondary data use requires REB approval
- **Data Retention & Destruction** – may be retained indefinitely for discovery, access, archival in accordance with REB-approved protocol

RISK MATRIX - EXTREME

From: **Sensitive data Toolkit for Researchers-** Part 2: Human Participant Research Data Risk Matrix
https://doi.org/10.5281/zenodo.4060159

Extreme

Data acquired through agreement with custodian, barring further use, retention; identifiers remain or (re)identification possible/probable; contains confidential, private, sensitive data; subjects are vulnerable in context of research and would be harmed in case of breach

- **Consent** – confidentiality will be maintained for as long as data exists; data will not be shared beyond research team
- **Collection** – researcher may know identities of participants; will have promised confidentiality through informed consent; methods of collection wide-ranging, may involve direct interaction with participants; direct identifiers may or may not be collected, but indirect identifiers collected sufficient to render participants identifiable (triangulation of data potential)
- **Analysis** – direct identifiers replace asap with linking code and separated physically and/or electronically from master list; consent forms with identifiers will be stored separately from research data; indirect identifiers should be coded, if possible; data shall only be access by research team members, as described in approved protocol; access/analysis shall only occur in a secure environment
- **Storage** – all data shall be stored on a centralized, standalone computer/site that is both password protected and encrypted, in appropriate secure location; backup consistent with risk level
- **Mobility/Sharing** – data restricted to a centralized, stand-alone computer/site that is password protected and encrypted; files should not be copied or shared; access shall be restricted to authorized individuals explicitly identified in REB protocol and involved least number possible
- **Deposit and Access** – data should not be deposited anywhere beyond the direct storage and access needs of the research team
- **Data Retention & Destruction** – data must be destroyed at earliest opportunity, in accordance with REB-approved protocol