Berenica Vejvoda, Research Data Librarian

bvejvoda duwindsor.ca

May 8, 2025

Best Practices for Managing your Research Data





DOCUMENTING YOUR DATA

SECURE STORAGE AND BACKUP

DATA SHARING AND ARCHIVING

What are research data?

Data used as primary sources to support research/scientific inquiry/creative practice

Almost anything that is analyzable!

- Text
- Numbers
- Symbols
- Images
- Films
- Video
- Sound recordings
- Pictorial reproductions
- Drawings
- Diagrams
- Algorithms



Is your data at risk?

- In a few years will <u>you remember</u> all the details of your research (e.g., variable names, response options, coding)?
- Would <u>someone else understand</u> it? (e.g., collaborator, supervisor, new researcher)
- Fast forward 5 years: can you <u>find and re-use</u> your data?
- What happens when you graduate, move positions, retire?
- If you lost your data, how much work would that be? Can you even recollect it (e.g., special events)???

HELP! LOST DATA FOR MY PHD

Alright, so, my Google Drive wasn't syncing 1,296 files. I saw that you can re-sync by deleting the DriveFS folder in the local Google folder. Did that and files started to upload, great. HOWEVER, all the previous files that didn't sync are now gone. I basically lost 8 months of work. I can't simply redo it because I work with mice and tumors that take a long time to work with.

I cut and pasted the old cache file to a new folder but the file is so small it can't possibly have all my data.

There's also the fact that I cut and pasted instead of using the "recover" tool in the recycle bin. Could that be an issue?

And no, there's nothing in my drive trash bin or anything like that. The location address just simply does not exist.

I found that someone downloaded their new Google drive then placed in the old cache then downloaded everything that was lost, then got in the new cache and simply added the lost files in. Again, the old cache is so tiny though.

I'm using drill disk right now. Do you think that will work?



r/talesfromtechsupport • 10 yr. ago meatwad75892

92

Lost a flash drive, lost a thesis.

Normally when I have a story to tell, it's one of my hundreds of clients from work doing something. But not the other day. This one is actually an old friend from high school that called me up one day around 7pm. I thought we were going to catch up on old times or something, but what she actually called about was her losing her thesis. (She is a doctoral candidate, going into education/teaching)

I walk her through letting me remote into her computer, and she tells me what happened. **She stored her thesis on 1 flash drive** and that's it. No backup copies, no syncing files to the cloud, nothing. Her entire thesis was on a flash drive. And she **lost it.** So I'm digging around the Autorecover folder on her computer, to no avail. It's not cached anywhere, and all I can really tell her is that it's gone and I have no way to help her. In complete denial, she begs me to look harder. So I start digging around folders that obviously won't have it just to appease her. (system32 et al) I tell her again that it's lost and that she should not have been storing it on only a flash drive.

Pretty clean-cut, boring case of user irresponsibility, but it still blows my mind every time it happens.





Predicted probability that the data were extant (either "shared" or "exist")

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K. J., Moore, J.-S., Renaut, S., & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, *24*(1), 94–97. https://doi.org/10.1016/j.cub.2013.11.014

What is RDM?

- Research data management (RDM)—processes applied through the lifecycle of a research project to guide the collection, documentation, storage, sharing and preservation of research data
- RDM is essential throughout the data lifecycle—from data creation, processing, analysis, preservation, storage, to sharing and reuse (where appropriate), at which point the cycle begins again

Why should you care about RDM?

- RDM enables researchers to organize, store, access, reuse and build upon digital research data
- RDM is essential to Canadian researchers' capacity to securely preserve and use their research data throughout their research projects, reuse their data over the course of their careers and, when appropriate, share their data
- RDM is a component of research excellence, strong RDM practices support researchers in achieving scientific rigor and enable collaboration in their fields

Fictitious Case Study for SoE

- Graduate student at UWindsor researching watersheds
- Collects terabytes of raster and vector images using GIS
 - Image files (raster)
 - Software GIS shapefiles (ArcGIS, vector)
 - Spreadsheets
 - Text files



Fictitious Case Study for SoE

- Data stored in a few places
 - Image files (10+ TB) on lab computers and external HDs
 - Smaller files (20GB) on personal laptop and institutional OneDrive
- Data is not documented consistently
- Data is not backed up
- Data was not archived
- Data is by request access



Fictitious Case Study for SoE

- There is a flood in SoE and the external hard drives are damaged leading to significant data loss
- Faced with recollecting data over the course of many years
- Finish and publish based on incomplete dataset



Better Practices



Make a plan for your data



Organize and document your data consistently



Store and backup your data securely



Prepare your data for archiving and sharing

Make a plan for your data (DMP)

- Create a DMP (Data Management Plan) from as soon as you have your project idea (avoid problems before they happen)
- DMP is your plan for how you will create, store, organize, document, secure, preserve and share your research data
- Both for <u>active</u> phases of research as well as for addressing <u>post</u> <u>completion</u>
- Living document (adapt and change through out project)
- Requirements from grant funding agencies (Tri Agencies, 2021)

Making a plan for your data

https://dmp-pgd.ca/



U Warning:

May 15, 2025 12:00-14:30 ET/9:00-11:30 PT // le 15 mai 2025 12:00-14:00 HE/9:00-11:30 HP

On May 15 there will be a complete outage of the DMP Assistant service. This time will be used for essential work and completion of the migration.

Le 15 mai, il y aura une interruption complète du service de l'Assistant PGD. Ce temps sera utilisé pour les travaux essentiels et l'achèvement de la migration.

My Dashboard

The table below lists the plans that you have created, and that have been shared with you by others. You can edit, share, download, make a copy, or remove these plans at any time.

roject Title	¢	Template	¢	Edited	•	Role	Test	Visibility	Shared	

- Bilingual tool
- Available to all researchers in Canada
- Setup as a wizard to step through all questions for managing your research data
- Templates for different types of research
- DMPs vary across disciplines, types of data, different methodologies
- Exportable plans
- Make as complete as possible (Tri Agencies don't evaluate for grant acceptance right now)
- 200+ samples (search by funder, discipline, location)

Data Management Plan for Longitudinal Monitorig of Freshwater Ecosystems under Climate Change Stressors

Project Details	Contributors	Plan overview	Write Plan	Research Outputs	Share	Download
expand all colla	ose all			0/24		
Data Collecti	on (0 / 5)					+
Documentat	ion and Metada	ta (0 / 3)				+
Storage and	Backup (0 / 2)					+
Preservation	(0 / 3)					+
Data Sharing	and Reuse (0 /	4)				+
Ethics and Le	egal Compliance	e (0 / 3)				+
Responsibilit	ies and Resourc	ces (0 / 4)				+

Templates

Digital Research Alliance of Canada: Alliance Template for Arts-Based Research

Data Collection

What types of data will you create and/or collect? What methods, arts-based and otherwise, will you use?

Example Answer:

Drawings, songs, poems, films, short stories, performances, interactive installations, and social experiences facilitated by artists are examples of data. Data on artistic processes can include documentation of techniques, stages, and contexts of artistic creation, and the physical materials (e.g., paints, textiles, found objects) and tools (e.g., pencils, the body, musical instruments) used to create artwork. Other types of data are audio recordings of interviews, transcripts, photographs, videos, field notes, historical documents, social media posts, statistical spreadsheets, and computer code.

Guidance:

Artwork is a prominent type of data in ABR that is commonly used as content for analysis and interpretation. Artworks that exist as, or are documented in, image, audio, video, text, and other types of digital files facilitate research data management. The same applies to preparatory, supplemental, and discarded artworks made in the creation of a principal one. Research findings you create in the form of artwork can be treated as data if you will make them available for researchers, artists, and/or the public to use as data. Information about artistic processes can also be data. Read more on artwork and artistic processes as data at <u>Kultur II Group</u> and <u>Jisc</u>.

Do you plan to use datasets published by others? Where will you collect them from?

Guidance:

Researchers and artists can publish their data for others to reuse. Research data repositories and government agencies are sources of published data (e.g., <u>Federated Research Data Repository</u>, <u>Statistics Canada</u>). Your university may have its own research data repository. Academic journals may host published data as supplementary material connected to their articles. If you need help finding resources for published data, contact your institution's library or reach out to the Portage DMP Coordinator at <u>dmp-assistant@tech.alliancecan.ca</u>.

How will you digitally document artwork, artistic processes, and other non-digital data? What conditions, hardware, software, and skills will you need?

Guidance:

Non-digital data should be digitized when possible. Digitization is needed for many reasons, including returning artwork to participants, creating records of performances, and depositing data in a repository for reuse. When planning your documentation, consider what conditions (e.g., good lighting, sound dampening), hardware (e.g., microphone, smartphone), software (e.g., video editing program), and specialized skills (e.g., filming techniques, image-editing skills) you will need. High quality documentation will make your data more valuable to you and others.

What file formats will your data be created and/or collected in?

Guidance:

Open (i.e., non-proprietary) file formats are preferred when possible because they can be used by anyone, which helps ensure others can access and reuse your data in the future. However, proprietary file formats may be necessary for certain arts-based methods because they have special capabilities for creating and editing images, audio, video, and text. If you use proprietary file formats, try to select industry-standard formats (i.e., those widely used by a given community) or those you can convert to open ones. <u>UK Data Service</u> provides a table of recommended and acceptable file formats for various types of data.

Templates

Digital Research Alliance of Canada: Alliance Template for Neuroimaging in the Neurosciences - Phase

Data Collection

Describe the types of data, and potential data sources, to be acquired during the course of your study.

Guidance:

Please explain, in particular:

- What type of neuroimaging modalities will be used to acquire data in this study? Ex: MRI, EEG.
- What other types of data will be acquired in this study? Ex: behavioural, biological sample.
- Approximately how many participants does the study plan to acquire images from?

Documentation and Metadata

How will you document your methods in order to support reproducibility?

Guidance:

For fellow researchers, a write-up of your methods is indispensable for supporting the reproducibility of a study. In preparation for publishing, consider creating an online document or folder (e.g. openneuro, github, zenodo, osf) where your project methods can be gathered/prepared. If appropriate, provide a link to that space here.

Storage and Backup

How and where will your data be stored and backed up during your research project?

Guidance:

Planning how research data will be stored and backed up throughout and beyond a research project is critical in ensuring data security and integrity. Appropriate storage and backup not only helps protect research data from catastrophic losses (due to hardware and software failures, viruses, hackers, natural disasters, human error, etc.), but also facilitates appropriate access by current and future researchers. You may need to encrypt your data to ensure it is not accessible by those outside the project. For more information, see the University of Waterloo's <u>Guideline for researchers on securing research participants' data</u>.

Please provide URL(s) to any data storage sites. If your data are subject to strict rules governing human subjects and anonymity, then you may need an on-premise solution installed on your institution's server.

What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.)?

Preservation

How will you store and retain your data after the active phase of data collection? For how long will you need to keep your data?

Guidance:

Choices about data preservation will depend on the potential for reuse and long-term significance of the data, as well as whether you have obligations to funders or collaborators to either retain or destroy data, and what resources will be required to ensure it remains usable in the future. The need to preserve data in the short-term (i.e. for peer-verification purposes) or long-term (for data of lasting value) will influence the choice of data repository or archive. Tools such as <u>DataCite's repository finder tool</u> and <u>re3data.org</u> are useful for finding an appropriate repository for your data.

Template for UWindsor

University of Windsor	uwindsor.ca	🗹 Contact	
--------------------------	-------------	-----------	--

Templates

If you wish to add an organizational template for a Data Management Plan, use the 'create template' button. You can create more than one template if desired e.g. one for researchers and one for PhD students. Your template will be presented to users within your organization when no funder templates apply. If you want to add questions to funder templates use the 'customize template' options below.

Own Templates Custom	izable Templates			
Own Template	25			
Template Name	Description	Status	Edited Date	\$
University of Windsor		Published 🕜	22-03-2022	Actions -
	l	1	1	I

Public DMPs

University of Windsor

<u>uwindsor.ca</u> <u>Contact</u>

Public DMPs

Public DMPs are plans created using the DMP Assistant service and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.

Search				
Project Title	Template 🗘	Organization	Owner	Download
Canada Excellence Research Chair in Nano-Optical Biosensing and Molecular Diagnostics	McMaster - NSERC Pilot DMP	McMaster University	Niko Hildebrandt	<mark>ل</mark> ـ
Optimizing quinoa production systems for Ontario - A physiology-based approach to improved agronomics	University of Guelph Agri-Food Funding Template	University of Guelph	Hugh Earl	۲
Avec toi sous un autre toit!	Portage Template for Mixed Methods (Surveys & Qualitative Research)	Centre hospitalier de l'Université de Montréal	Louise Allard	<mark>ب</mark>
Development of high-yielding soybean cultivars with diverse genetic background, disease resistance and seed quality	Ontario Agri-Food Innovation Alliance Template	University of Guelph	lstvan Rajcan	<mark>ب</mark>
Regional approach to advance integrated management tools to minimize disease loss from tar spot on corn	Ontario Agri-Food Innovation Alliance Template	University of Guelph	David Hooker	لم لم
Early detection of Acute Kidney InjurD	Portage Template	Other Organisation	First Name Surname	لم لم
FFTARI-2023-5- Enabling no-till corn and soybean production in high-residue environments via improved planter components	University of Guelph Agri-Food Funding Template (Alliance/Food from Thought)	University of Guelph	Joshua Nasielski	<mark>ک</mark>

DMP Tool (U.S.)

Public DMPs

https://dmptool.org/

>>>> DMP Tool

Public Plans 🛛 Funder Requirements About 🌐

Sign Up

Login

Public Plans

Public plans are plans created using the DMP Tool service and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.

Find a plan	Plans (1238)	Sor	t by: Featured ~
Search	FEATURED		
		DMP ID:	10.48321/D1WK99
Clear Search	Development of Cutibacterium-specific immunoassays	Creation date:	06-01-2023
Funder (124) +	to identify true Cutibacterium acnes infections.	Language:	English (US)
Institution (300)	William Mccoy iv, Washington University in St. Louis (wustl.edu)		

Public DMPs

Public DMPs are plans created using the DMPonline service and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.

Project title	Template 🔷	Organisation	Owner	Download
Onderzoek BMW - leerlijn Data Science - update en voorbeeld	UMC Utrecht DMP with DPIA V.3.0	UMC Utrecht	John Meeuwsen	POF
Assessing & Enhancing Energy Efficiency of HVAC Systems in Saudi Arabia's Residential Buildings	University of Manchester Generic Template	University of Manchester	Mustafa Alobaid	PDF
Unhealthy Food and Convenience Trends in Households with Dependents	University of Nottingham generic Data Management Plan	University of Nottingham	Simon Welham	PDF
Monitoring Earth's Evolution and Tectonics, WP11 - IPSES - Italian Platform for Solid Earth Science, Activity 11.9a - Services and interoperability layers for Iistributing earthquake faulting data in 4D	DCC Template	Other	Roberto Basili	PDF
Pottery Production at Roman Corbridge: its character, scale and significance	DCC Template	Newcastle University	Ludovica	

DMP ONLINE (UK)

<u>https://dmponline.dcc.ac.uk/public_plans?page=ALL</u>

- Projects can mean the creation of hundreds of data files
- A year from now can you identify variable labels?
- Would you be clear on the <u>methodology</u> used about <u>when/where/how the data were</u> <u>collected</u>?
- Do you know where you stored the scripting code needed to re-run your analysis?

- · Short descriptive file names
- · Simple file hierarchy
- Easier to navigate and locate files
- Document file names for all team members
- Be consistent

- Make organizing files and folders a habit so you can readily know where things go
- File organization can include:
 - By project
 - By researcher
 - By [experiment/study]
 - By date
 - Combination of above such as year->project

- Naming files descriptively, consistently and in a standardized way will help make your data easier to understand, share and archive
- File naming: include date, project name, short description, researcher initials, version number, and other metadata (e.g., location)
 - 2024_01_31_GLIER_Watershed_Guidelines_UWstaff_JK_v01.docx
 - Make sure it is YYYYMMDD versus MMDDYYYY as computers sort by YYYYMMD
 - GLIER (Great Lakes Institute for Environmental Research) [short unique identifier]
 - Use _ [underscore] as delimiter; avoid spaces between words and these special characters: & , * % # * ()! @ \$ ^ ~ ` { } [] ? < > [different systems use different characters differently]

- Once you have your organizational schema/naming system, document it in a text <u>READme.txt file</u> so that the entire research team knows
- **README**: A simple text document (.txt) which describes project information, folder hierarchy, file organization and description of essential file contents
- Data dictionary: A document for structured tabular data which describes variable names and labels (including definitions and exact questions asked, response options[1=agree], record layout [structure of data, specifying the order, type, and length of each field]
- **Codebooks**: Often includes a data dictionary but also includes survey methodology (weighting, missing values, data quality, cell suppression, sampling method, data reliability issues)
- Statistics Canada (aka, 'Documentation and User Guide'') E.g., <u>2021 Census Individuals PUMF User</u> <u>Guide.pdf</u>

2021 Census Public Use Microdata File (PUMF)

Individuals File Documentation and User guide





Statistics Statistique Canada Canada



- Core elements of a README.txt file, include:
 - Researcher contact information
 - Use license for your data
 - Goal of your research
 - · Data collection methods (sample, instruments)
 - File structure
 - File naming conventions
 - Quality assurance work (e.g. how you made sure the data is good quality)

- A file format is a way to encode information within a computer file so that it can be recognized by an application and accessed
- This is indicated by the file name extension (e.g. .txt, .doc, .jpg, .mov).
- Allows a computer to recognize that a document contains text or that a file should be processed as a video
- File formats are essential for data storage and preservation.
- Software and data storage technology changes quickly, and files can easily become obsolete or difficult to access
- Recommended that data files are copied to new media every 2-5 years, especially if technology changes

- Using open file formats (e.g., .Rdata, .Rds versus .sav, .dta .sas)
- Open file formats can increase ability to open and read files in the future
- Open file formats can be used by anyone
- Standardized, well documented, commonly used, uncompressed
- Makes data usable and accessible to more researchers sooner
- Opensource software communities can ensure that data stored in these file formats remain accessible over the long term



- Proprietary File Formats work only with software provided by the company that produces them
- Not freely available
- When software is no longer supported, files in that format are typically unreadable
- Some research disciplines and industries treat a specific proprietary file format as a *de facto* standard which you may wish to follow
- Industry may be more convenient to use
- Can archive both open and proprietary
- Recommended:
 - Databases: XML, CSV
 - E-Books: EPUB
 - Images: JPG, PNG, PDF, TIFF, BMP
 - Sound: MP3, FLAC
 - Text: TXT, CSV, PDF/A, ASCII, UTF-8
 - Video: MPG, MOV, AVI
 - Spreadsheets: CSV
 - Medical Images: DICOM

- Each file type such as text, images, or sound has many file formats available (e.g., JPEG, TIFF, PNG, GIF, PSD, EPS, AI, RAW)
- File quality is a large part of the file format decision
- High resolution will be larger than
- · lower quality file formats
- Cost of storage space
- **Disseminating** the file to others (convenience)

Secure storage and backup



Data Loss

Device theft Device loss Hardware & software failures Accidental damage Natural disasters Human error



IT Security

Viruses, malware, ransomware



Redisclosure risk

Secure storage and backup **CYBER SECURITY NEWS** • 2 MIN READ **Ransomware Gang Claims to Have Stolen 430 GB of Internal Data in Stanford University Data Breach**

ALICIA HOPE · NOVEMBER 9, 2023

Stanford University is investigating an alleged data breach after the Akira ransomware gang claimed it breached the institution and exfiltrated 430 GB of data, including private information and confidential documents.

Secure active storage and backup

- 3-2-1 Back Up Rule (at least 3 copies!)
 - 3 copies of your data (at minimum)
 - 2 copies on-hand (easily accessible) on different systems (e.g., internal/external HD and secure campus drive) [working and backup copy]
 - 1 copy in another location (off site with trusted service provider e.g., institutional OneDrive)

Secure active storage and backup

- Include metadata and README files in your data backups
- Campuses will have IT solutions to help you
- UWindsor: <u>https://leddy.uwindsor.ca/rdm</u> <u>/storage-security</u>

Research Data Management



Secure active storage and backup Active Data storage options for the University of Windsor are listed here. Active storage refers to your needs for research data storage while you are actively working with your research data during the course of your research project. These options are distinct from depositing and archiving your data for long term preservation for posterity and so that your data can be discovered and potentially reused by others. For data preservation options, please see the section for Archiving Data. Also, please be advised that many of the options listed are not ideal for very sensitive research data that is not de-identified or sufficiently anonymized.

Digital Research Alliance (the Alliance) Advanced Research Computing

The Digital Research Alliance of Canada (the Alliance) is a national governmental organization supporting advanced research computing (services formerly provided by Compute Canada). The Alliance provides researchers with access to high performance computing (HPC). Usually this is a good option for researchers working with larger datasets which exceed the limits of desktop computing. The Alliance therefore provides large amounts of storage for researchers and data is securely transferred to the Alliance ARC systems. The systems are not specifically set up for handling sensitive data, but support staff can provide advice. By default ITB is assigned but more storage space can be assigned through research competitions. All storage options can be found on the Alliance ARC storage website. Servers are all hosted in Canada. Globus is the preferred protocol for transferring data to and between Alliance ARC systems. File storage can also be accessed using SFTP, SCP, or other network protocols. Using ARC storage requires high technical complexity (e.g., working in Linux). Faculty members at any Canadian University can make a CCDB account by following the instructions here. Research staff and graduate students accounts can be sponsored by faculty researchers. For questions, please contact the Alliance contact page.

OneDrive

University of Windsor's institutional OneDrive access is hosted locally on institutional servers and is a viable option for secure active storage. As of April 7, 2025, space available to faculty was reduced to **20GB of space** - Faculty have been granted an extension to September 1, 2025 to address the reductions in available storage through Microsoft. Users who may require more than the allotted quota for work, academic, or research purposes can submit a request to IT Services for review.

Nextcloud (Digital Research Alliance of Canada)

The Digital Research Alliance of Canada (the Alliance) is a national governmental organization supporting advanced research computing (services formerly provided by Compute Canada). The Alliance provides Compute Canada Nextcloud, an online file syncing and storage solution for all advanced research computing (ARC) users. The Nextcloud service is aimed at users with relatively small datasets (**100GB** are assigned by default). If there is a need, researchers may be able to request more storage space. Medium risk data must be manually encrypted before being uploaded to Compute Canada Nextcloud. Nextcloud is hosted at SFU and is backed up daily. Users can share files with other users, create and send password-protected public links, and let others upload files to the private cloud. Nextcloud access is automatically provided to all Alliance ARC users, if you have a CCDB account you can login with it. If you do not have a CCDB account, faculty members at any Canadia University can apply for a CCDB account by following the instructions here. Faculty researchers can then sponsor research staff and graduate students accounts as well. Unlike advanced research computing, NextCloud does not require high technical skill, making this a suitable option for those who prefer web portal acces (versus command line). Desktop syncing is possible for Mac, Windows, and Linux and Android and IOS mobile apps.

Ontario Library Research Cloud (OLRC) has been developed by the Ontario Council of University Libraries (including the University of Windsor) as an Ontario-based, highly secure cloud solution. The servers are located in Ontario University controlled data centres and the data travels over a private VLAN. Currently lacks a scalable user interface. For extremely sensitive projects with infrequent data transfer and access, the OLRC may be an option.

Local Secure Networks

Departmental networked storage (Novell etc.) may be a secure viable option, but check with your administrator on security and access protections.

Local storage

Your local hard drive may be the most convenient option for primary storage of most extremely sensitive data at present. Make sure you have external backup copies and in more than one location! Also, encrypt and password protect files to minimize security risk.

REDCap

REDCap (Research Electronic Data Capture) is a secure web application for building and managing online surveys and databases. If you are collecting survey data, then the University of Windsor's locally-installed instance of RECap is a very good option for secure local storage and is highly secure. REDCap is more of a data collection tool than an active storage platform. After data collection is complete, researchers are encouraged to move their data from REDCap. REDCap is free for all VE-SPARK Health Institute Members.

Secure storage and backup

- Free cloud services such as Dropbox and Google Drive are **not recommended** and most ethics boards and granting agencies prohibit them from research data sharing or storage
- Use a secure environment to share selected research data with individuals and/or groups [data transfer is important too!]
- Keep data secure during all phases of your research project
- Ensure that all computer hard drives (and other devices such as USB keys and flash drives) are **encrypted to prevent unauthorized access** to information



Secure storage and backup

- Encryption process of transforming information so that it is only readable to a person with the correct authorization
- On personal machine use full disk encryption (iOS: FileVault; Windows: Bitlocker or ''device encryption'')
- Especially important if you are collecting and/or working with sensitive data that would create harm if released openly – e.g., personally identifiable information and personal health information
- Use strong passwords (strong [vs too simple], unique, secret, all devices, monitor changes in password)



Data sharing: open and free data sharing supports as well as restricted use: Data archiving: ensures your research data remains accessible and useable (to you and others) for the long term

transparency, accessibility, reproducibility and replicability, new collaboration, maximizing impact and visibility of research, sharing with community partners, and prevents over researching communities

Data as ''public good" – Federal funding requirements

Research Data Storage Continuum

Figure 1. Research data storage spectrum (the Alliance RDM WG, 2020). © All Rights

In March 2021, the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada, and the Social Sciences and Humanities Research Council of Canada <u>announced</u> the launch of the <u>Tri-Agency Research Data</u> <u>Management Policy</u> (the Policy).

Search Canada.ca

MENU 🗸

- <u>Canada.ca</u> > <u>Science and innovation</u> > <u>Science.gc.ca</u> > <u>Interagency research funding</u> > <u>Policies and Guidelines</u>
- > <u>Research Data Management</u>

Tri-Agency Research Data Management Policy

Institutional strategies: By March 1, 2023, research institutions subject to this requirement must post their RDM strategies and notify the agencies when they have been completed.

Data management plans: By spring 2022, the agencies will identify the initial set of funding opportunities subject to the DMP requirement. The agencies will pilot the DMP requirement in targeted funding opportunities before this date.

Data deposit: After reviewing the institutional strategies and in line with the readiness of the Canadian research community, the agencies will phase in the deposit requirement.

• **CIHR-funded researchers:** Since January 1, 2008, recipients of CIHR funding have had to comply with the limited data deposit requirements included in the Tri-Agency Open Access Policy on Publications. They must continue to comply with these requirements, which are specific to bioinformatics, atomic, and molecular coordinate data.

CIHR: Network Grants in Skin Health and Muscular Dystrophy; Data Science for Equity; Virtual Care/Digital Health Team Grants (launch Fall 2022, application deadline Winter 2023)

NSERC: Subatomic Physics Discovery Grants (Individual and Project, 2023)

SSHRC: Partnership Grants Phase 2 (2023)

The citation advantage of linking publications to research data

Giovanni Colavizza, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker, Barbara McGillivray

Published: April 22, 2020 • https://doi.org/10.1371/journal.pone.0230416

Data sharing and

archiving

Article	Authors	Metrics	Comments	Media Coverage Peer Review			Download PDF	
*							Print	Share
Abstract	Abstr	act						
Introduction	Efforte	lll-	the anal and reproduc	ible are increasingly re	eflected by journal		Check for	or updates
<u>к</u>		•	Efforts to make re by journal policie statements.	esearch results ope s encouraging or m	n and reproducible and ating authors to	are incr provide	reasingly refle e data availa	ected bility

In 2018 93.7% of 21,793 PLOS articles and 88.2% of 31,956 BMC articles had data availability statements. Data availability statements containing a link to data in a repository—rather than being available on request—are a fraction of the total. In 2017 and 2018, 20.8% of PLOS publications and 12.2% of BMC publications provided DAS containing a link to data in a repository.

- Publications in PLOS and BMC with open data had up to a 25% higher citation impact vs. those not sharing data
- Journals with higher impact factors are more likely to have data sharing policies

24,757

View

860

Share

Journals encouraging data sharing and include data availability statements:

PLOS ONE

Introduction

Minimal Data Set Definition

Acceptable Data Sharing Methods

Acceptable Data Access Restrictions

Unacceptable Data Access Restrictions

FAQs

PLOS Data Advisory Board

Data Availability

The following policy applies to all PLOS journals, unless otherwise noted.

Introduction

PLOS journals require authors to make all data necessary to replicate their study's findings publicly available without restriction at the time of publication. When specific legal or ethical restrictions prohibit public sharing of a data set, authors must indicate how others may obtain access to the data.

advanced search

• Journals encouraging data sharing and include data availability statements:

SPRINGER NATURE Springer Nature Research data policy

doi Awaiting DOI

Туре

Registry

Journal publisher

Policy

Description

At Springer Nature we advance discovery by publishing trusted research, supporting the development of new ideas and championing open science. We also aim to facilitate compliance with research funder and institution requirements to share data. To help accomplish this we have established a standard research data policy for our journals, based on transparency around supporting data. This policy applies to all datasets that are necessary to interpret and replicate the conclusions reported in a research article. This policy mandates the use of a data availability statement, and endorses (but does not mandate) submission into an appropriate domain-specific repository except under certain conditions. A large number of our journals already support this policy, including Nature Portfolio, BMC and many Springer and Palgrave titles. They are in the process of implementing this policy across the remainder of our portfolio, in stages. For information on a journal's specific policies, please consult the journal submission guidelines. Please note: This policy supersedes Springer Nature's (legacy) tiered research data policy, which viewed here: can be https://www.springernature.com/gp/authors/research-data-policy/research-data-policy-types.

Homepage

https://www.springernature.com/gp/authors/research-data-policy

Journals encouraging data sharing and include data availability statements:

Data Access Statement Type	Description	Example Text	Published Example (if available)
Openly available data	Data is openly available in a public repository. Access is provided by listing a link to the data repository in the published paper.	All data files are available from [nome of public data repository] at [web link that includes DOI and reference number]. NOTE: The web link is generated by the repository when data is denovited.	Research Integrity and Peer Review
		unpromeso.	Avian Research
Secondary analysis of existing data	Data is openly available and researchers can list the location of the original dataset as well as any new data generated from the secondary data analysis.	This study is a re-analysis of existing data that are publicly available from [web link that includes DOI and reference number]. Further documentation about data processing is available from [web link that Includes DOI and reference number].	Enviornmental Health
Ethical constraints (Restricted data is available on request)	Data is restricted for ethical reasons and either unavailable, or available upon request to bona fide researchers who meet a pre- specified criteria.	Due to ethical restrictions, data are available upon request from [contact information].	BMC Psychology
			<u>PloS One</u>
Legal constraints (Restricted data is available on request)	Data is restricted for legal reasons and either unavailable, or available upon request to bona fide researchers who meet a pre-specified criteria.	Due to legal restriction, all relevant data cannot be made publicly available. Data are available from the [institution] for all interested researchers who meet the criteria for access to confidential data. Data can be made available upon request to [contact information]. NOTE: ideally the contact information is not to a specific person in order to provide long-term access.	<u>Plos One</u>
Data available through a third party	Data cannot be distributed due to limitation related to rights (e.g., data is under licence from a commercial provider)	The data used in this analysis is owned by [name of owner] and the authors do not have permission to make it publicly available; however, interested researchers would be able to access these data by request for permission addressed to [contact information].	PLoS Dne
Data is available on a web page	Data is publicly accessible and provided on a public website (rather than a public data repository)	The data used in this study are available from [provide link to website]. Other supplemental data used in this study are also publicly available and the sources have been appropriately cited in the text.	<u>PloS One</u>
Data is embargoed	Data is unavailable for a stated period of time but will be made openly available in the future. State a specific end date (10 October 2020) of the embargo period rather than a time period (6 months).	Data will be available at [provide link] after a [list time period] embargo and will be available on [dote] to allow for commercialisation of research findings. This data is subject to an embargo and will be released on [dote].	
Data is within article or as supplementary materials	Data is presented in the paper as tables, figures or by other methods. Alternately, it is provided as appendices or supplementary files.	All relevant data are within the paper and its supporting information files.	BMC Medicine
			Big Data Analytics
No new data created	Data was not generated or collected in the course of the study.	Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.	

- Depositing data into a research data repository is a good way to ensure data will be accessible over the long term:
 - Assignment of DOIs
 - Long term storage infrastructure
 - Sufficient metadata (study and variable-level) to allow for discovery for reuse
 - Control access (embargoes, sensitive data encryption)

Domain-specific – e.g., genomic, astrophysics

General repositories –e.g., Alliance's Borealis & FRDR

Code repositories –e.g., Github and Gitlab

Use a repository finder such as DataCite or re3data

https://borealisdata.ca/datave rse/windsor

 <u>https://borealisdata.ca/d</u> <u>ataverse/windsor</u>

1	to	4 of	4 F	iles
---	----	------	-----	------

GEC ELEM BOUND.7z

7Z Archive - 37.9 KB - Oct 22, 2019 - 2 Downloads MD5: e6489173cdaeddd20a5c7f31064e712a Greater Essex County School Board elementary school catchment boundary files in GIS format.

	\square

GEC_SEC_BOUND.7z

7Z Archive - 19.8 KB - Oct 22, 2019 - 0 Downloads MD5: 3df2d35bdfa36551af7c243dfb284793 Greater Essex County School Board secondary school catchment boundary files in GIS format.

WEC_ELEM_BOUND.7z

7Z Archive - 23.9 KB - Oct 22, 2019 - 0 Downloads MD5: 72ea3200c6f565bdfdc513870429ba33 Windsor Essex Catholic District School Board's elementary school catchment boundary files in GIS format.

WEC_SEC_BOUND.7z

7Z Archive - 12.6 KB - Oct 22, 2019 - 0 Downloads MD5: eb3427e0b10286988d1e62c79d7632d6 Windsor Essex Catholic District School Board's secondary school catchment boundary

files in GIS format.

https://www.frdrdfdr.ca/repo/ Intended for long term preservation of larger datasets. **1TB of repository storage** is available to all faculty members at Canadian post-secondary institutions. Some key features include:

- Fast and efficient data upload and download for large datasets through Globus File Transfer, as well as web browser transfers for smaller datasets
- DOI registration for datasets provides a persistent identifier that can be used for data citation
- Support for multiple authentication providers, including Compute Canada, ORCID, and a growing list of Canadian postsecondary institutions
- Optional embargoes on data and metadata records
- Allows multiple people to collaborate on a submission
- Secure repository storage for a minimum of 10 years after deposit
- Regular backups and geographically distributed storage for research data
- Additional storage may be provided upon request.

https://www.re3data.org/

	REGISTRY OF RESEARCH DAT	ta.	Drg
--	--------------------------	-----	-----

<u>https://www.re3data.org</u>
/

Browse by country

<u>https://www.re3data.org</u>
//

https://leddy.uwindsor.ca/rdm

Research Data Management

Training and Events

https://leddy.uwindsor.ca/rdm /training-events

Training and Events Key Service Areas Training Research Data and Statistical The Digital Research Alliance of Canada offers a range of training materials - everything from one page guides to online training modules Services - that span the research data life cycle. Archives & Special Collections University of Windsor Webinars, Workshops and Events Digital Scholarship and Preservation Depositing and Sharing Data for Preservation in Borealis (2) slides) (Leddy Library and the Office of Research and Integrity Services, April Copyright 24, 2025) **Research Data Management** Building a Data Management Plan (recording) (2 slides) (co-sponsored by Leddy Library and the Office of Research and Integrity Services, April 10, 2025) University of Windsor's RDM Research Data Managemeent and Text Data Mining in JupyterHub with Newspapers (co-sponsored by Leddy Library, SHARCNET, and Strategy Brock University, February, 2023) University of Windsor RDM RDM Terminology Strategy Committees Alliance RDM Glossary Data Management Plans **Brief Guides** Metadata & Organizing Data Brief Guide – Create an Effective Data Management Plan Storage and Security Brief Guide – Data Curation Brief Guide – Data Management Plan Protecting Sensitive Data Brief Guide – Dataverse Metadata Archiving Data Brief Guide – FRDR. Brief Guide – Institutional Strategies Sharing Data Brief Guide – Research Data Management Tri-Agency Research Data Primers Management Policy Primer – Curation **Training and Events** Primer – Data Management Plan RDM & TDM in Primer – Dataverse JupyterHub with Primer – FRDR Newspapers Primer – Institutional Strategies Primer – Research Data Management Scholarship and Publishing Research Data Management 101 Connect with your library Module 1: Background and Learning Objectives Module 2: What is RDM (Research Data Management)? Module 3: Canadian Policy Review Module 4: Steps Towards Good Research Data Management Research Data Repositories 101 These modules will help users learn about the role that research data repositories play in good RDM, and in meeting the requirements of the Tri-Agency RDM Policy. Module 1: Background and Learning Objectives Module 2: What are Research Data Repositories? Module 3: Types of Research Data Repositories Module 4: TRUST and FAIR Principles for Research Data Repositories · Module 5: Research Data Policies Landscape: Funders', Publishers', and Institutional Module 6: Glossary of Terms Module 7: Test Your Knowledge!