

Berenica Vejvoda, Research Data Librarian, Leddy Library April 24, 2025

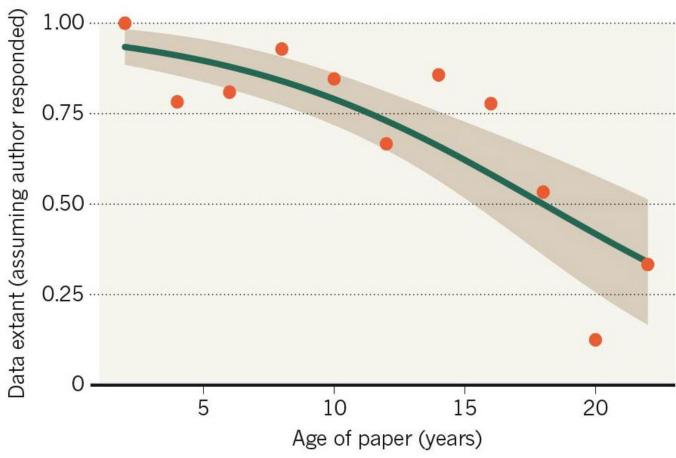
Why share data?

- Research data are a valuable resource, usually requiring much time and money to be produced. Many data have significant value beyond usage for the original research.
- Funders and journals increasingly require that data be published in a trustworthy repository as part of a well-developed data management plan.
- Depositing research data facilitates data reuse across and within disciplines.
- The ability to find and re-use data is increasingly important for verifying published research findings and supporting new research.
- Depositing your data is also a good way to ensure data remain accessible beyond the life of the study for which they were collected.

Why share data?

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



Gibney, E., Van Noorden, R. Scientists losing data at a rapid rate. *Nature* (2013). https://doi.org/10.1038/nature.2013.14416

Why should I deposit my research data?

Funder requirements

Journal and publisher requirements

Citation Impact

Archival and preservation

A key component of the Tri-Agency
Research Data Management (RDM)
Policy, launched in March 2021, is
the requirement mandating that
researchers deposit digital research
data, metadata, and code
supporting their published findings
in a recognized repository.

While there will be a deposit mandate, public sharing depends on ethical, legal, or commercial considerations and follows FAIR principles (Findable, Accessible, Interoperable, Reusable).

Tri-Agency RDM Policy requirements



Institutional strategies

Each post-secondary institution and research hospital eligible to administer CIHR, NSERC or SSHRC funds is required to create an institutional RDM strategy, notify the agencies when it has been completed, and post it publicly.



Data
Management
Plans (DMPs)

For certain funding opportunities, the agencies will require DMPs to be submitted to the appropriate agency at time of application, as outlined in the call for proposals.



Data deposit

Grant recipients are required to deposit into a digital repository all digital research data, metadata and code that directly support the research conclusions in journal publications and pre-prints that arise from agency-supported research.

Funder requirements

Data Deposit is part of the <u>Tri-Agency</u> Research Data Management Policy.

Tri-Agency RDM Policy – Indigenous research



Institutional strategies

Recognizing that data created in the context of research by and with First Nations, Métis, and Inuit communities, collectives and organizations will be managed according to principles developed and approved by those communities, communities and organizations, and in partnership with them; recognizing distinctions-based approaches.



Data
Management
Plans (DMPs)

For research conducted by and with First Nations, Métis and Inuit communities, collectives and organizations, DMPs must be codeveloped with those communities, collectives and organizations, in accordance with RDM principles or DMP formats that they accept.



Data deposit

For research conducted by and with First Nations, Métis and Inuit communities, collectives and organizations, those communities, collectives or organizations will guide and ultimately determine how the data are collected, used and preserved, and have the right to repatriate the data. This could result in exceptions to the data deposit requirement.

Why are the agencies requiring data deposit?

- Ensure secure data retention
- Make data accessible (when appropriate)



- Promote research integrity (reproducibility, trust)
- Facilitate data reuse: open / controlled-access data reduce effort duplication and promote collaboration

Implementation timeline



Institutional strategies

By **March 1, 2023**, each postsecondary institution and research hospital must post their RDM strategy.



Data Management Plans (DMPs)

By **Spring 2022**, the agencies will identify the initial set of funding opportunities subject to the DMP requirement.



Data deposit

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

Implementation of the data deposit requirement

Community engagements since 2023 on topics including:

- desirable characteristics of repositories
- FAIR principles / metadata standards
- 'as open as possible, as secured as necessary'
- deposit and controlled access to sensitive data
- Indigenous data sovereignty
- compliance monitoring / responsible research assessment
- education, training and infrastructure needs

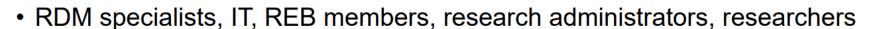
Engagements to date

- Digital Research Alliance of Canada
 - RDM leadership
 - RDM Network of Experts: Council of Chairs, SDEG, DMPEG
 - Controlled Access Management (CAM) Initiative
- Service providers
 - Data repositories: FRDR, Borealis, Dryad, OSF
 - Other services: PIDs (CARL, ORCID), monitoring (OpenAlex, DataSeer)
- Funding partners
 - Provincial/national funding agencies: FRQ, RNS, Horizon Europe, UKRI, NSF, NWO, SNSF
 - Philanthropies: Open Research Funders Group, Gates Foundation, Wellcome Trust



Engagements to date

- Other partner organizations (OCSA, CRDCN, CRKN)
- Indigenous RDM and data sovereignty
 - Indigenous Leadership Circle in Research
 - SSHRC Indigenous Strategy Team
 - Canadian Network of Indigenous Research Administrators



- In person engagements leveraging conferences in Calgary and Halifax in May 2024
- On request by institutions (e.g., Queen's, Windsor, Saskatchewan)
- Internal engagements (SRCR, Research Security and Impact teams, TGMS)



Examples of key takeaways

- Key concepts and expectations should be clarified in the policy text / FAQs.
 (What data must be deposited? What is a recognized repository?)
- The international policy landscape has changed since the RDM policy was developed → 'as open as possible, as closed/secured as necessary'.
- Sensitive data are a challenge, and suitable Canadian repositories will take time to develop. Process for accessing sensitive data must be clearly stated.
- Data curation is important for reusability, but capacity varies among institutions.
- There is a need for investments in education, training, support & infrastructure.

Anticipated data deposit implementation timeline

Approach to implementation

(ongoing)

Community engagement

(2023-25)

What We Heard report

(first half of 2025)

Text for public comments

(mid 2025)



Anticipated data deposit implementation timeline

Communications with lead-time

(2nd half of 2025)

Requirement comes into effect

(anticipated 2026)

First data deposited

(~2 years later)

Monitoring and compliance (RRA)

(2028 and beyond)

Many journals are starting to require data sharing or at least data availability statements, including:

- PLOS: journals.plos.org/plosone/s/data-availability
- Nature: nature.com/nature-portfolio/editorial-policies/reporting-standards
- NEJM: <u>nejm.org/about-nejm/editorial-policies</u>
- Journals with higher impact factors are more likely to have data sharing policies.

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.

When submitting a manuscript, authors must provide a Data Availability Statement describing compliance with PLOS' data policy. If the article is accepted for publication, the Data Availability Statement will be published as part of the article.

Acceptable data sharing methods are listed below, accompanied by guidance for authors as to what must be included in their Data Availability Statement and how to follow best practices in research reporting.

PLOS believes that sharing data fosters scientific progress. Data availability allows and facilitates:

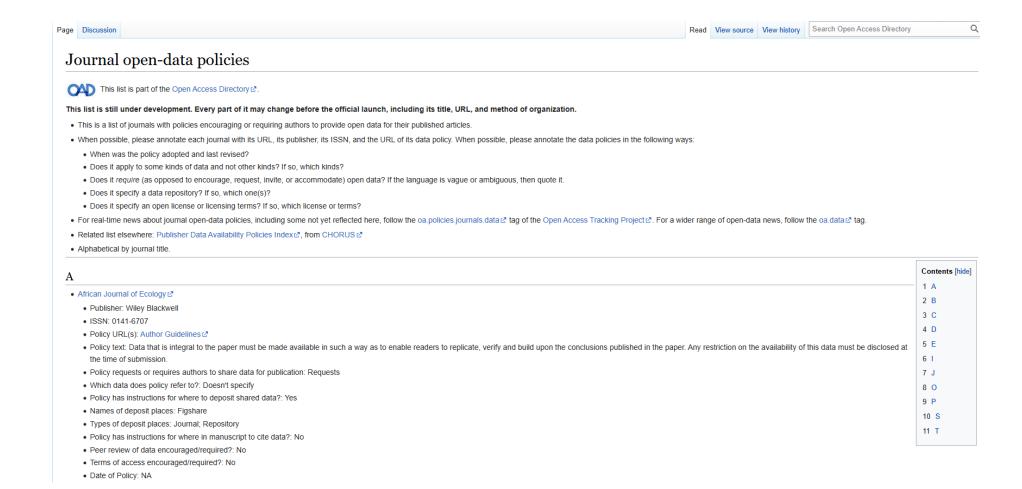
- Validation, replication, reanalysis, new analysis, reinterpretation or inclusion into meta-analyses;
- Reproducibility of research;
- > Efforts to ensure data are archived, increasing the value of the investment made in funding scientific research;
- > Reduction of the burden on authors in preserving and finding old data, and managing data access requests;
- > Citation and linking of research data and their associated articles, enhancing visibility and ensuring recognition for authors, data producers and curators.

Publication is conditional on compliance with this policy. If restrictions on access to data come to light after publication, we reserve the right to post a Correction, an Editorial Expression of Concern, contact the authors' institutions and funders, or, in extreme cases, retract the publication.



PLOS strongly recommends sharing data in a repository whenever possible. Data repositories improve discoverability and accessibility, ensure long-term preservation, and lead to increased attention for the research.

PLOS journals highlight repository use with our Accessible Data icon, an experimental feature that appears on any research article with a link to selected repositories in its Data Availability Statement.



https://jordproject.wordpress.com/project-data/social-science-journals-that-have-a-research-data-policy/

jord

ACCESS TO RESEARCH DATA: ADDRESSING THE PROBLEM THROUGH JOURNAL DATA SHARING POLICIES

CONTACTS PROJECT DATA PROJECT OUTLINE REPORTS AND ARTICLE

PROJECT INFORMATION LITERATURE REVIEW STAKEHOLDER CONSULTATION DATA COLLECTION

DATA REPOSITORIES DATA SHARING POLICY

Social Science Journals that have a research data policy

These journals were taken from the 2011 Thomson Reuters' Journal citation reports. They include the journals from the top 100 (most cited) and bottom 100 least cited) ranked by impact factor.

Journal Title	ISSN	URL
American Economic Journal- Macroeconomics	1945- 7707	http://www.aeaweb.org/aej/mac/index.php
Journal Of Marketing	0022- 2429	http://www.marketingpower.com/AboutAMA/Pages/AMA%20Publications/AMA%20
Archives Of General Psychiatry	0003- 990X	http://archpsyc.jamanetwork.com/journal.aspx
American Journal Of Psychiatry	0002- 953X	http://www.appi.org/SearchCenter/Pages/Journal.aspx?ItemId=AJP
American Psychologist	0003- 066X	http://www.apa.org/pubs/journals/amp/index.aspx
American Journal Of Public Health	0090- 0036	http://ajph.aphapublications.org/userimages/ContentEditor/1318438422261/Instruc
Annual Review Of	1548-	http://www.annualreviews.org/page/authors/author-instructions/preparing/supmat

1 Discipline	Journal/Publisher	licy Overview						K	
2 Life Sciences	PLOS Biology / PLOS ONE	Requires public da	ta availability	as a cond	dition of p	ublication.			
3 Life Sciences	GigaScience	Mandates open data and code availability in structured repositories.							
4 Life Sciences	Scientific Data (Nature)	Requires full data deposition in recognized repositories.							
5 Life Sciences	Animal Conservation	Requires DNA sequence data to be deposited in EMBL/GenBank/DDBJ databases.							
6 Life Sciences	Aquatic Microbial Ecology	Mandates deposition of nucleotide or protein sequences in GenBank, EMBL, or DDBJ databases.							
7 Life Sciences	Evolution	Requires DNA data in GenBank and phylogenetic data in TreeBase.							
8 Life Sciences	ISME Journal	Requires all non-article data to be publicly available in databases.							
9 Health Sciences	ВМЈ	Requires data sharing statements for clinical trials; encourages open data.							
10 Health Sciences	BMC Medicine / BMC Genomics	Require data availability statements; many mandate deposit.							
11 Health Sciences	Journal of Exposure Science and Env	Journal of Exposure Science and Envir Mandates data sharing policy aligned with Nature standards.							
12 Psychology	Journal of Open Psychology Data	Publishes only arti	cles with ope	n dataset	·S.				
13 Psychology	Judgment and Decision Making	Requires open sha	ring of data u	nderlying	analysis.				
14 Psychology	Comprehensive Results in Social Psy	Comprehensive Results in Social Psycl Requires authors to share raw data upon publication.							
15 Psychology	Journal of Applied Psychology	Journal of Applied Psychology Requires data transparency and variable disclosure.							
16 Social Sciences	Taylor & Francis (Social Science Jour	Taylor & Francis (Social Science Journa Tiered policies; many journals require or encourage deposit.							
17 Social Sciences	SAGE (Social Science Journals)	SAGE (Social Science Journals) Implements tiered data sharing policies.							
18 Social Sciences	American Economic Journal: Macroe	American Economic Journal: Macroecc Encourages data sharing; policy may vary.							
19 Social Sciences	Journal of Marketing	Encourages data sh	aring; policy	may vary.					
20 Social Sciences	American Journal of Public Health	Encourages data sh	aring; policy	may vary.					
21 Social Sciences	Annual Review of Sociology	Encourages data sh	aring; policy	may vary.					
22 Social Sciences	Global Environmental Change	Encourages data sh	aring; policy	may vary.					
23 Multidisciplinary	Springer Nature	Requires data avai	lability staten	nents; ma	andates de	posit for m	any.		
24 Multidisciplinary	Elsevier	Requires data avai	lability; encou	urages or	mandates	deposit.			
25 Multidisciplinary	Wiley	Requires data avai	lability; policy	y varies.					
26 Multidisciplinary	F1000Research	Requires open dat	a and method	transpar	ency.				
27 Multidisciplinary	eLife	Open science mod	el; data and o	ode shou	ld be avail	able.			
28 Multidisciplinary	Science (AAAS)	Requires public da	ta availability	for verifi	ication and	reuse.			



University of Toronto RDM website: https://onesearch.libr ary.utoronto.ca/sites/ default/public/rdm/da taaccessstatements.p df

	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 [name 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 deposited.	Research Integrity and Peer Review	
				<u>Avian Research</u>	
	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 Plas One	
	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.	pon request to bona fide researchers who meet a pre-specified available. Data are available from the [institution] for all interested		
	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].	<u>PLoS One</u>	
	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 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.		

Citation impact

The Citation Advantage of Sharing Data

- Increased Citations: Studies have shown that articles linking to datasets in repositories can experience a citation boost of up to 25% compared to those without such links. arXiv
- Open Science Practices: Engaging in open science practices, including data sharing, correlates with a 4.3% average increase in citations. <u>arXiv</u>

Citation impact

- •Enhanced Discoverability: Proper data citation improves the findability and accessibility of datasets, aligning with FAIR principles. <u>Labor Dynamics Institute</u>
- •Academic Recognition: Citing datasets allows researchers to receive credit for their work, contributing to their academic reputation. PMC+1PLOS+1

Citation impact

Implementing Data Citation

- Persistent Identifiers: Assigning DOIs to datasets ensures consistent and reliable citation.
- Repository Selection: Choosing the right repository (disciplinary, generalist, or institutional) is crucial for maximizing the impact and accessibility of your data.

Archival and preservation

Even in the absence of funder requirements, you may choose to keep your research data because:

- Your data can support the integrity of your research by providing evidence of your research methods and findings
- Your data may be unique or irreplaceable
- You may want to reuse your data in the future
- You may want to share your data with other researchers

What to deposit?

- Data as a digital package
 - A curated set of data files and supporting documentation prepared to ensure long-term accessibility and usability. Unlike your working files, a data package includes only the most relevant and finalized materials. Preparing a data package often involves organizing, renaming, and documenting files in a clear and consistent way. The main objective is to make the dataset understandable and usable by others outside your immediate research team.
- Research output
- Identifiers
- Metadata (study and file level)

Preparing for depositing your data

- Verifying that all required files and accompanying documentation are present
- Creating or updating documentation to explain how the data can be accessed and reused, including license information
- Organizing files and providing clear descriptions
- Identifying and correcting code errors or other quality issues
- Assessing datasets for potential privacy or disclosure risks
- Converting files into formats suitable for long-term preservation
- Enhancing and reviewing metadata for accuracy and completeness

Types of data repositories

There are three primary categories of data repositories:

- Disciplinary repositories serve specific research domains or data types. These platforms
 often have established guidelines for file formats, documentation, and metadata
 standards. To locate a disciplinary repository, you can consult with colleagues in your
 field, check journal submission guidelines, or explore re3data, a global registry of
 research data repositories. York U list.
- Multidisciplinary or generalist repositories accept a wide range of data across all disciplines. These repositories are flexible in scope and include platforms such as FRDR, Dryad, Zenodo, and figshare.
- Institutional repositories are managed by universities or research institutions and typically support data produced by their affiliated researchers. For example, University of Windsor Borealis, is the University of Windsor's institutional research data repository. It is available to researchers conducting work at, or supported by, the University of Windsor.

Where should I deposit my research data?

Data repositories vary in the services and features they provide. When choosing where to deposit your dataset, consider the following factors:

- Funder and publisher requirements: Some funding agencies, academic journals, or publishers may mandate or recommend specific repositories for data deposit. It's important to verify these requirements early in your research process.
- Relevance to your discipline: Sharing data in a repository tailored to your field can increase its visibility and encourage reuse among peers. In some areas, depositing data in a disciplinary repository is a common or even expected practice.
- Persistent identifiers (PIDs): Many repositories assign a persistent identifier—such as a DOI—that makes citing and linking to your dataset easier. Be sure to check which type of identifier is provided and whether it aligns with your needs.

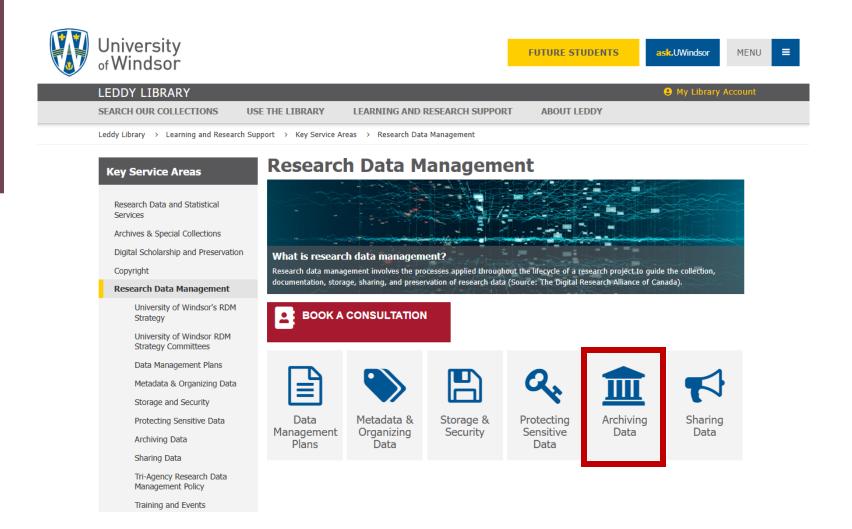
Where should I deposit my research data?

- •Access controls: Certain platforms allow data owners to place restrictions on who can access specific files or full datasets. This feature may enable you or the repository to review and approve requests before data is shared.
- •Licensing options: Repositories often let you choose from a range of licenses that dictate how others can use your data. Confirm that the available licenses, such as those from Creative Commons or Open Data Commons, suit your goals.
- •Costs: While many repositories are free, some charge fees for deposit or long-term storage. Evaluating repository options early in your project can help you plan for potential costs. The University of Windsor Borealis, for instance, is free for researchers affiliated with the University of Windsor.

Where should I deposit my research data?

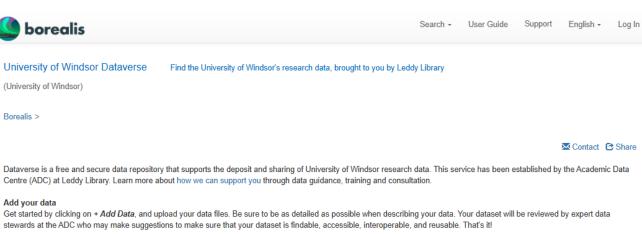
- •Retention and preservation policies: Repositories differ in how long they retain data and what steps they take to preserve it. Understanding these policies can help ensure your data remains accessible in the long term.
- •Ability to revise or update: Some repositories let you manage your dataset post-deposit—for example, by uploading new versions or updating metadata—while others limit this functionality.
- •Curation and support: Certain repositories provide curation services that enhance your dataset's value. This may include tasks like improving metadata, validating files, adding search-friendly keywords, or helping your data meet community standards. These efforts can improve discoverability and encourage reuse.

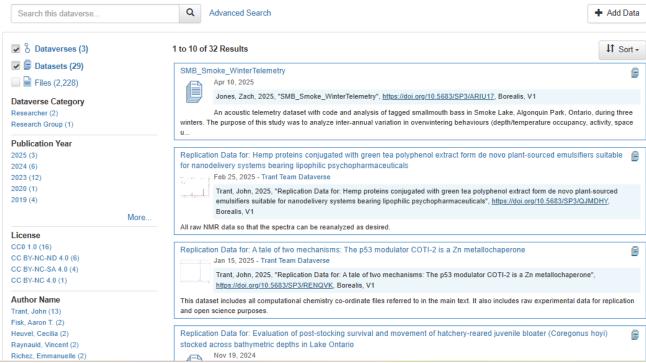
RDM Knowledge Portal



https://leddy.uwindsor.ca/key-service-areas/research-datamanagement

Borealis





Borealis

- This is a secure environment that conforms to industry best practices for maintaining data integrity and longevity.
- Data owners can retain complete control over who can access their data.
- Specific datasets can be made freely available, available only on request, or restricted to only certain IPs, to an individual account(s) or to specific groups.
- Borealis is widely used by academic institutions in Canada and the U.S.
 and datasets are indexed and searchable across institutions.
- Borealis is most suited for datasets up to 2GB.

Borealis

Key Features

- Free to University of Windsor researchers for storing (long term) and sharing research data (excluding sensitive/confidential data)
- Supports all file types, including documents, spreadsheets, stats files, images, audio, video, and zip files
- Each dataset receives a DOI and an auto-generated citation to support sharing and attribution
- Secure storage on U ofT servers with monthly integrity checks for long-term preservation
- Datasets are indexed for discovery via platforms like Google, Lunaris, and DataCite
- Flexible access controls allow open or restricted sharing at the file or dataset level

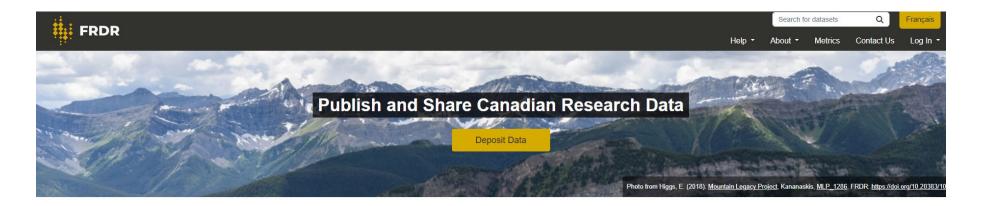
Why use Borealis?

- Secure storage (caveat not active storage or for sensitive data)
- Persistent identifiers
- Discoverability
- Controlled access
- Visualization and analysis tools
- Preservation

How is Borealis being used?

- 7,189 datasets published in 2023
- 443,963 data files deposited from over 21,678 datasets across 1,507 collections covering multiple disciplines
- 2,119 datasets have direct links to publications and other scholarly outputs. View open access research publications connected to data deposited in Borealis

Federated Research Data Repository (FRDR)



The Federated Research Data Repository (FRDR) is a bilingual publishing platform for sharing and preserving Canadian research data. It is a curated, general-purpose repository, custom built for large datasets.

Get started on FRDR



Create an Account

- Sign up or log in using your Institutional ID, Federation ID, ORCID, or a Globus ID
- · We'll verify your account
- More details in <u>Before</u>
 <u>Depositing</u>



Deposit Data

- Describe your dataset, select a license, and upload data files
- Optional: Invite collaborators, link to related research outputs
- · We'll reserve a DOI for your



Curate and Publish

- Internal review by a member of our curation team to optimize the FAIRness of your dataset
- Dataset published and DOI registered
- · Appraisal and selection for

Q

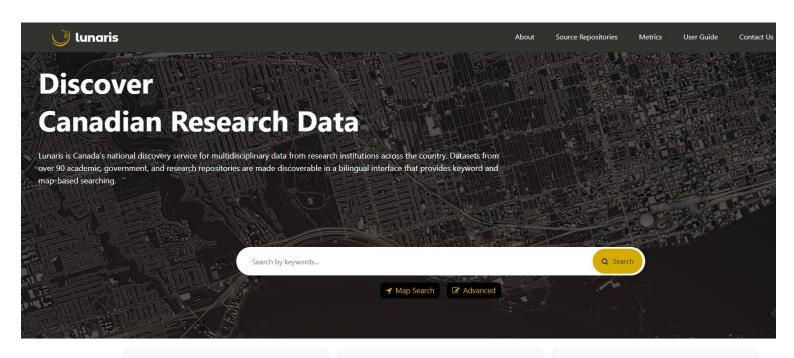
Discover and Download

- · Data indexed for discovery
- Cite and promote your dataset
- Anyone may use FRDR to find and download data
- Use <u>Lunaris</u> [™] to discover data

Federated Research Data Repository (FRDR)

- Fast and efficient data upload and download for large datasets through Globus File Transfer, as well as web browser transfers for smaller datasets
- 1TB of repository storage is available to Canadian faculty members (& dessignates)
- Robust documentation available in English and French
- 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
- Support for Special Collections by research groups with organizational branding and links to external web content
- A responsive support service with business hours coverage across Canada (support@frdr-dfdr.ca)
- Repository storage spread across multiple Compute Canada Federation hosting sites (geographically distributed storage)
- Secure repository storage for a minimum of 10 years after deposit
- Regular backups
- On-site system administration and ongoing hardware maintenance

Lunaris (discoverability layer for Borealis/FRDR)









Resources

- Primer Curation (Digital Research Alliance of Canada)
- Brief Guide Data Curation (Digital Research Alliance of Canada)
- <u>Data Curation Primers</u> (Data Curation Network)
- The DCN CURATE(D) Steps (Data Curation Network)
- Curating Datasets for Reproducibility (University of Victoria)
- University of Toronto README file creation guidance:
 - README for Data Deposit: Template this template can be used to create a README file to include in a data deposit.
 - <u>README for Data Deposit: Template Guide</u> this guide provides additional information and guidance about specific sections and fields in the README for Data Deposit Template.
- Registry of Research Data Repositories: https://www.re3data.org/