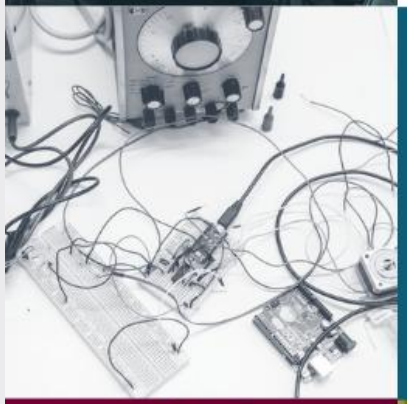
Data Deposit Bootcamp

Sherman Centre for Digital Scholarship

Tuesday, May 19, 2026 from 1:00pm - 4:00pm

Health Science Library eClassroom

Sherman Centre for Digital Scholarship



Data Deposit Bootcamp

Isaac Pratt, PhD and Danica Evering, MA

May 19, 2026

McMaster University Library

Land Acknowledgement

McMaster University is located on the traditional territories of the Mississauga and Haudenosaunee Nations. Settlers have responsibilities under the Silver Covenant Chain Wampum, part of the 1764 Treaty of Niagara.

Georgia Kirkos, “Campus Grounds Spring,” June 7, 2019, McMaster University, Hamilton, Ontario, Canada

https://mcmaster.assetbank.app/assetbank-mcmaster/action/viewAsset?id=22349&index=4&total=1000&view=viewSearchItem

Certificate Programs

The Sherman Centre for Digital Scholarship Certificate of Attendance

The Sherman Centre's certificate program recognizes attendance at our workshops. It complements degree training, supports the development of critical competencies in data analysis, research data management, and digital scholarship, and formalizes core skills fostered by our workshops.

Participants are invited to collect **seven** workshop points to receive a certificate of attendance. To verify your participation in today's workshop, we will provide a code and additional instructions at the end of the session.

You can learn more about the certificate program at **scds.ca/certificate-program**

The Canadian Certificate for Digital Humanities

This workshop is also eligible for the Canadian Certificate for Digital Humanities. To learn more about the certificate, visit **ccdhhn.ca**. You can also contact local liaison Alexis-Carlota Cochrane at scds@mcmaster.ca.

Outline

Goal Setting + Introductions [20 mins]

Data Deposit + Sharing Overview [15 mins]

Working Session [5 mins intro, 20 mins work time]

- **Documentation** – README Files, Data Dictionaries, Codebooks, and beyond.
- **File Preparation** – Anonymizing and organizing data.
- **Archiving** – Choosing what to keep and what formats to keep it in.
- **Findable + Reusable** – Licenses, metadata, and keywords.
- **Citation + Credit** – Publication DOIs, data sources, collaborators.

Next Steps – Submitting data for review [15 mins]

Research Data Management Services

PhD in Anatomy and Cell Biology, member of MREB - can help with data management for sensitive data

Isaac Pratt, PhD

MA in Media Studies, excited about data justice, community research, + connecting with curious disciplines!

Danica Evering, MA

scds.ca
scds@mcmaster.ca

A joint initiative of McMaster's Faculty of Humanities and McMaster University Libraries
McMaster University
Sherman Centre for Digital Scholarship

Who are you?

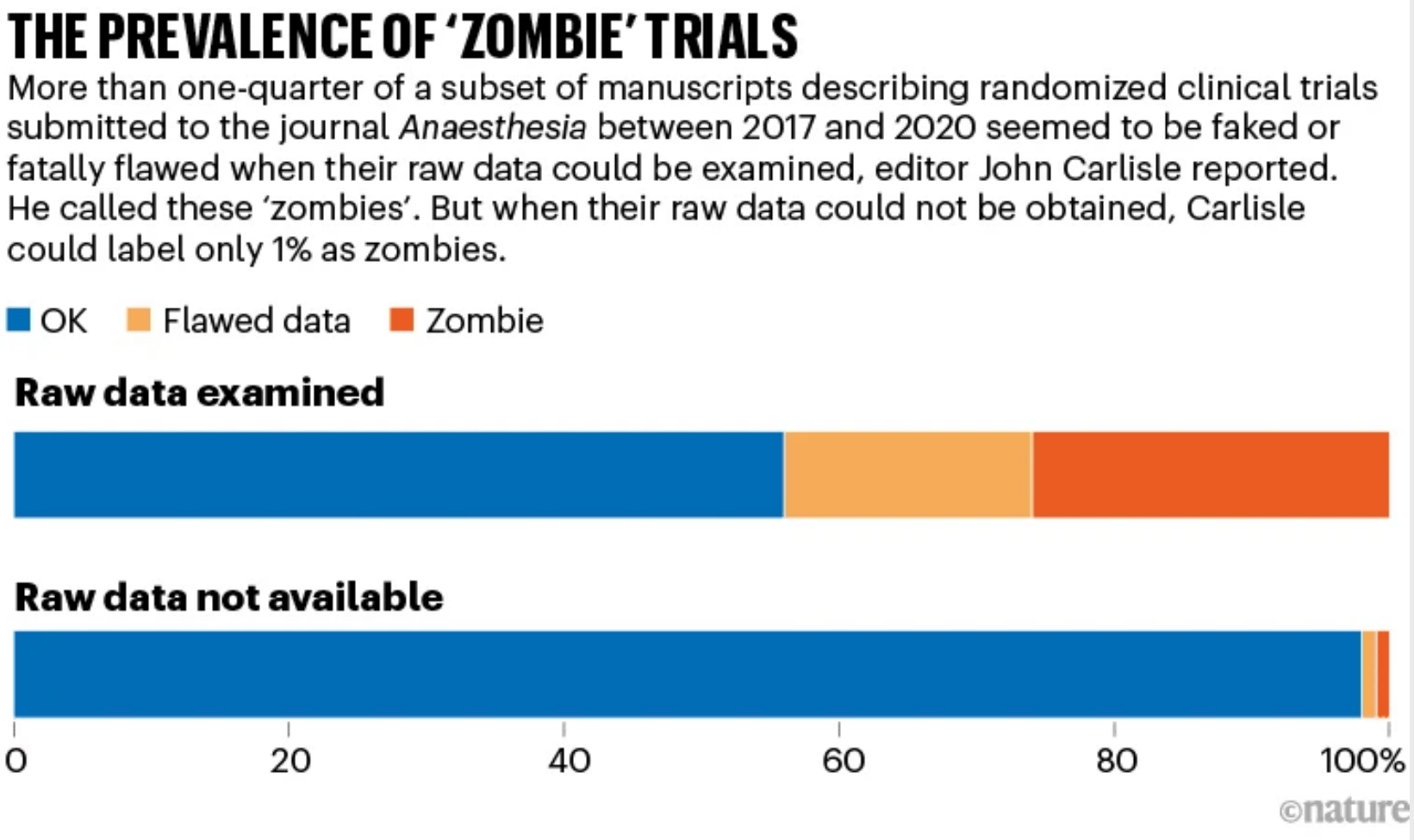
*Hands up if you are…*a **faculty** member? A **graduate student**? On a team as **research staff**? A **postdoc**? Some other role?

What's your goal for today?

Hands up if you are...depositing data for... a **journal publication**? For... your **thesis**? Or are you just here to **learn**?

Joel Friesen, "Group of zombies, shooting of the film Meat Market 3," 2006-06-13, Wikimedia Commons, CC-BY 2.0 https://commons.wikimedia.org/wiki/File:Groupofzombiesjoelf.jpg

‘Zombie’ Trials and Data Availability



Van Noorden, R. (2023). Medicine is plagued by untrustworthy clinical trials. How many studies are faked or flawed? Nature, 619(7970), 454–458. https://doi.org/10.1038/d41586-023-02299-w

Open Research

- Supports **reproducibility** and **replicability** – heightens confidence in results and supports research integrity
- Leads to **new collaborations** – potential for meta-analyses over a wider topic area.
- Research is often publicly funded – data as a “public good” that is accessible by journalists, other researchers, etc.
- Limits **over-researching communities** for qualitative studies. Share datasets with **community partners** – value added, supports long-lasting research relationships.

scds.ca
scds@mcmaster.ca

A joint initiative of McMaster’s Faculty of Humanities and McMaster University Libraries
McMaster University
Sherman Centre for Digital Scholarship

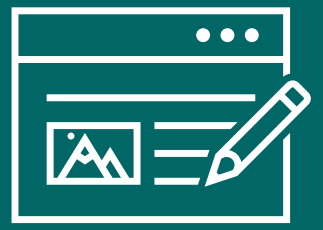
What can expanded open data sharing enable?

Researchers

- Increased citations + research impact
- Meet increased data requirements
- Increased collaborations and partnerships
- Avoid retractions
- Preservation of data

Research community

- More confidence in research results
- Increased ability to build on previous results
- Culture of reproducible research



Publishers

- Reproducible results
- More confidence in published results
- Fewer retractions

Funders

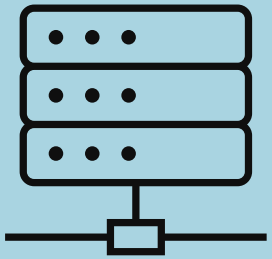
- Maximizes value of funding dollars
- Research excellence
- Alignment with international research community

Society

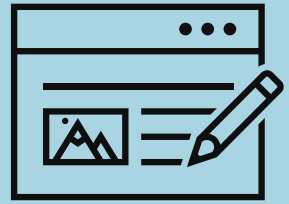
- Faster + greater benefits from research
- Increased public confidence
- Access outside of academia (journalists, NGOs, citizens)

Where does data live **long-term**?

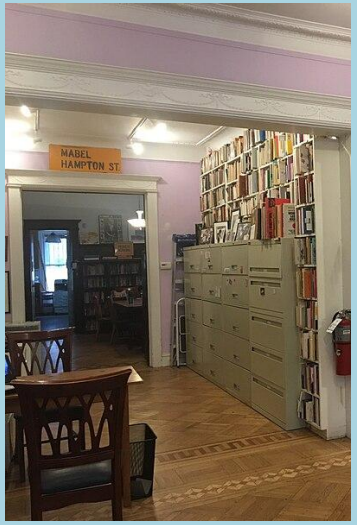
researcher/
research team

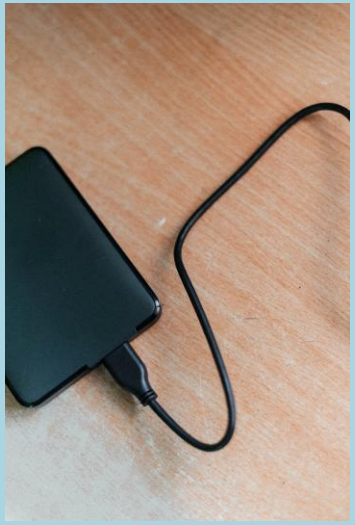


research data



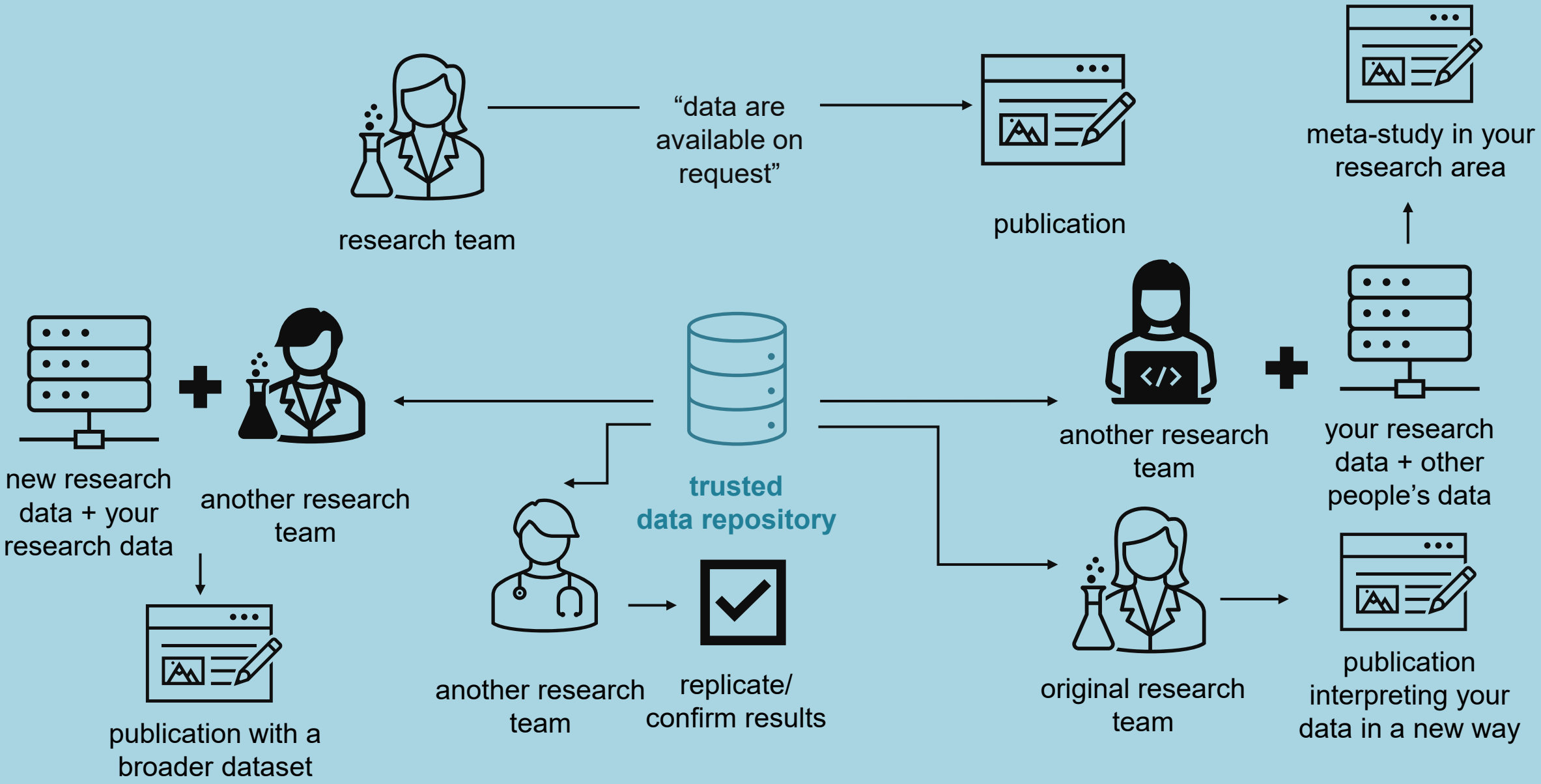
publication





?

What does data sharing look like?



Data Deposit requirements

Funders

- Tri-Agencies: “Grant recipients are required to deposit into a digital repository all digital **research data**, **metadata** and **code** that support the research conclusions in journal publications and pre-prints.” **Grant recipients not required to openly share data but provide appropriate access.**
- *Rollout TBD, expected late 2026*. CIHR has data related requirements currently in force.
- National Institutes of Health (NIH) – Data Management and Sharing Policy (2023)

Publishers

- Journals requiring data sharing or availability statements: BMJ, PLOS, Nature, NEJM, and Frontiers.

Why? *Ensuring data preservation, reproducibility and research integrity, reduce effort duplication and promote collaboration, return on public investment.*

Photo by Guillermo Ruiz on Unsplash.

Why share data? **Archival and Preservation**

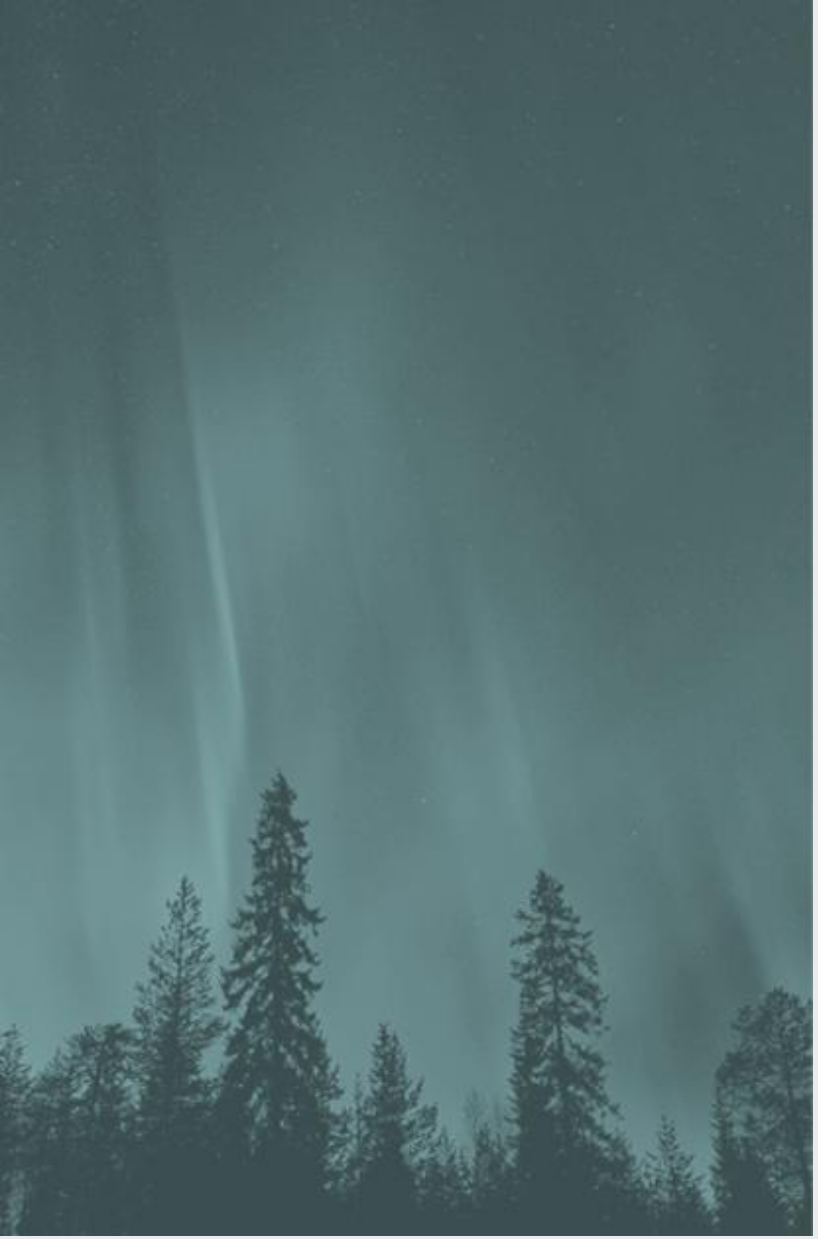
Sharing data in a research data repository is a good way to ensure that data will be accessible over the long term.

- Datasets are given **DOIs**
- Datasets are in platforms that have plans for long term storage and curation
- Datasets are published with enough documentation to enable re-use or re-analysis

A joint initiative of McMaster's Faculty of Humanities and McMaster University Libraries

McMaster University

Sherman Centre for Digital Scholarship



Where does data get deposited?

A **trusted data repository** is a web platform and storage space for researchers to deposit data sets associated with their research.

Repositories provide:

- long-term storage and access to research data beyond the life of a grant, research project, or individual careers
- discoverability and findability for datasets through features like indexing and DOIs
- easy-to-use shared platforms made for research

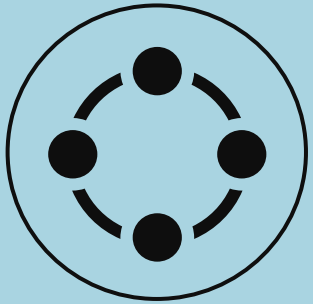
scds.ca
scds@mcmaster.ca

A joint initiative of McMaster's Faculty of Humanities and McMaster University Libraries
McMaster University
Sherman Centre for Digital Scholarship

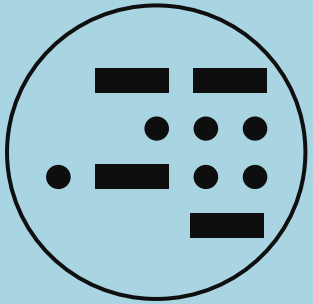
Data Repositories

Publishing data in a recognized data repository is the best way to share data. There are thousands of data repositories.

Domain Specific Repositories
Focus on certain types of data such as genomic information or astronomical information. https://www.nature.com/sdata/policies/repositories



General Repositories
Accept broader types of research data. *ex. McMaster Dataverse (part of Borealis) and Canada's Federated Research Data Repository (FRDR), Open Science Framework (OSF).*



Code Repositories
There are also code-specific repositories like Github, Gitlab, BitBucket, SourceForge

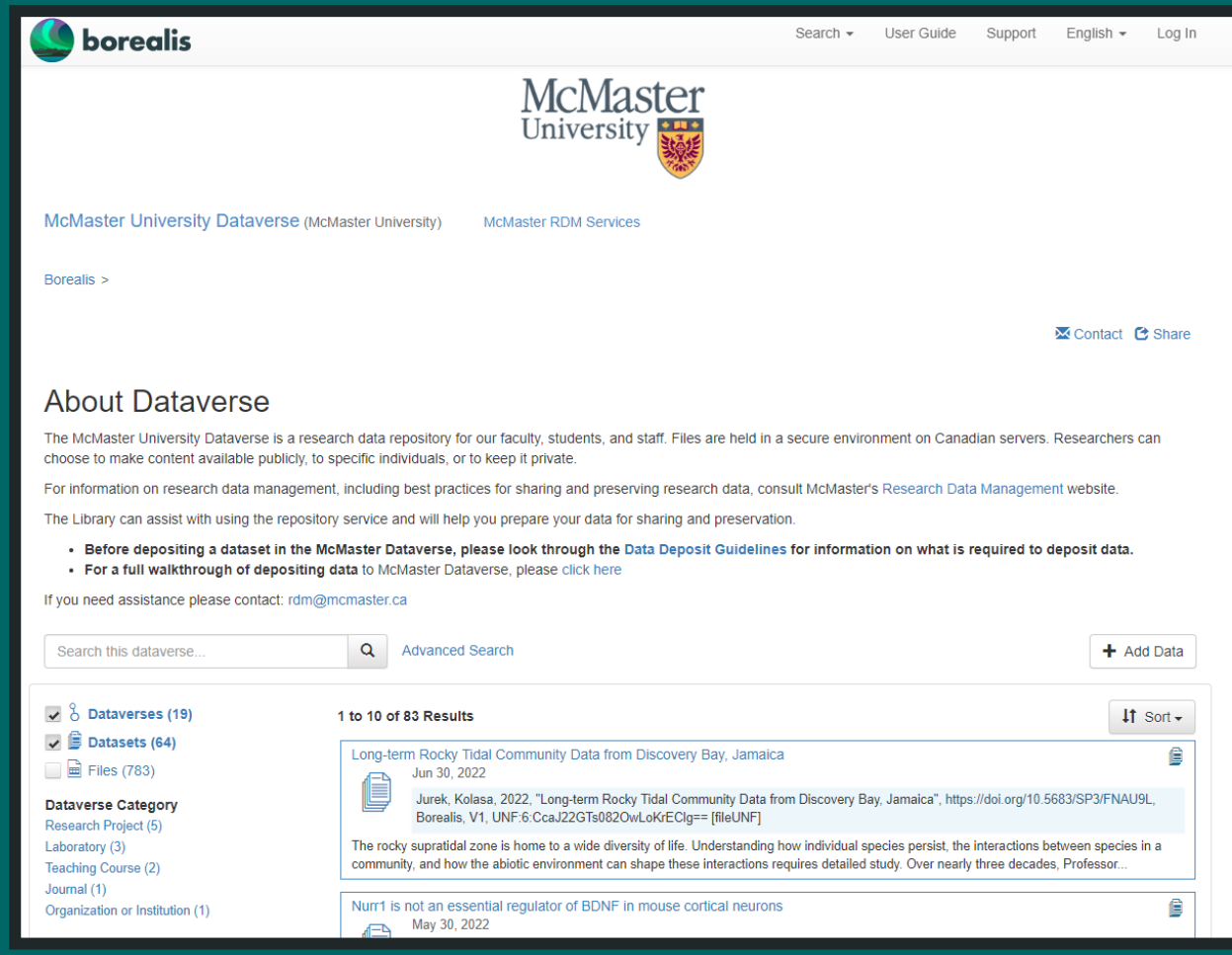
Controlled Access
For sensitive or qualitative data look at Vivli (Clinical Data), Qualitative Data Repository, Inter-university Consortium for Political and Social Research (ICPSR)

Where do colleagues in your field share their data?
Where do you look for secondary data?

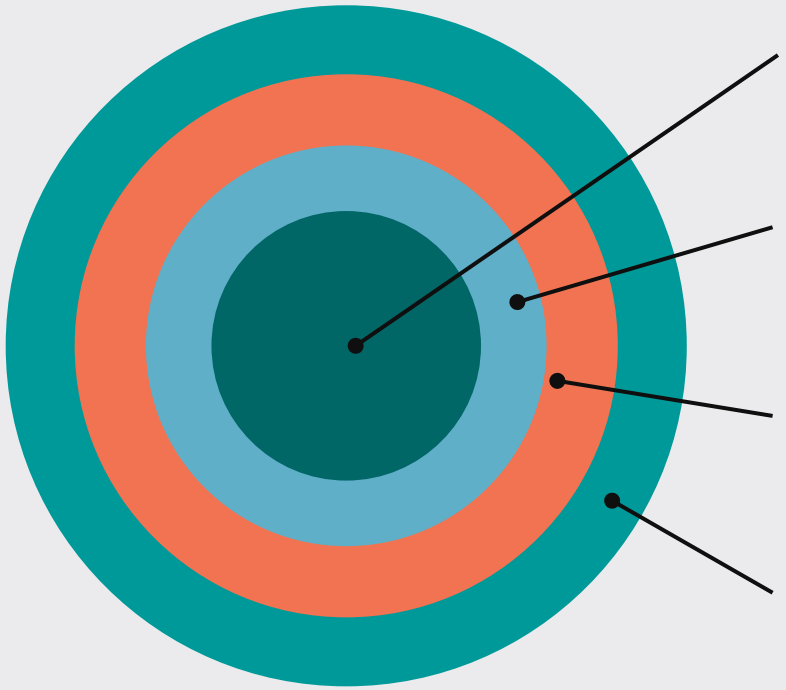
McMaster Dataverse

https://borealisdata.ca/dataverse/mcmaster

- McMaster's Institutional Data Repository is a home for research data created by McMaster researchers. *(Not recommended for sensitive data)*
- Provides basic data curation services
- Data is stewarded by professionals at McMaster
- Contains tools for tabular data exploration and analysis



Dataset as a digital package.



Research output (data/code): The data is surrounded by layers of information to make it FAIR

Identifiers: Persistent Unique Identifiers such as DOIs and ORCiDs help find, track, and cite data

Standards: Open standard file formats help others access and reuse data

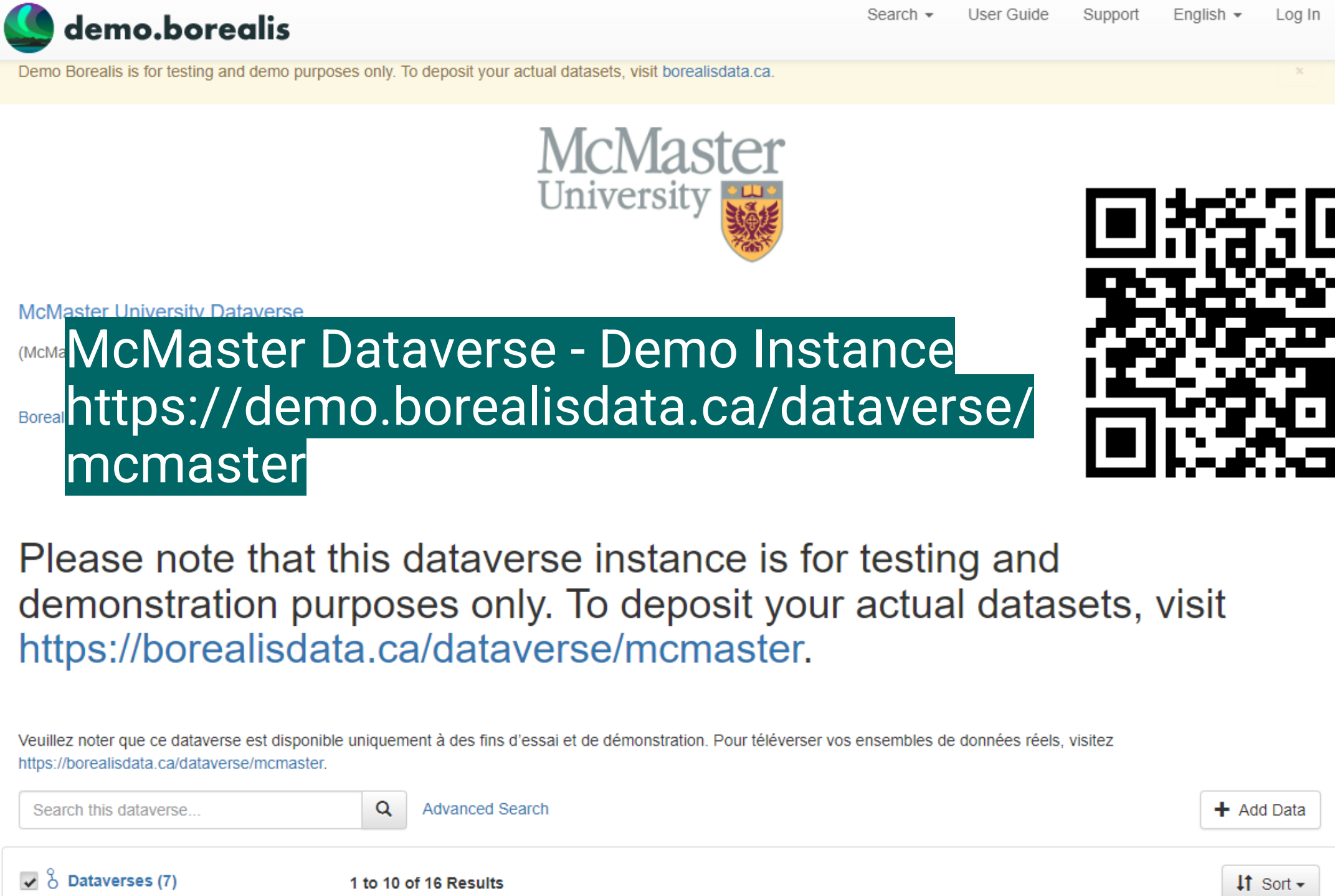
Metadata: Rich metadata and data documentation helps others find and understand datasets

scds.ca
scds@mcmaster.ca

A joint initiative of McMaster's Faculty of Humanities and McMaster University Libraries

McMaster University

Sherman Centre for Digital Scholarship


demo.borealis
Search
User Guide
Support
English
Log In
Demo Borealis is for testing and demo purposes only. To deposit your actual datasets, visit borealisdata.ca.
McMaster University
McMaster University Dataverse
(McMa
Boreal
McMaster Dataverse - Demo Instance
https://demo.borealisdata.ca/dataverse/mcmaster
Please note that this dataverse instance is for testing and demonstration purposes only. To deposit your actual datasets, visit https://borealisdata.ca/dataverse/mcmaster.
Veuillez noter que ce dataverse est disponible uniquement à des fins d'essai et de démonstration. Pour téléverser vos ensembles de données réels, visitez https://borealisdata.ca/dataverse/mcmaster.
Search this dataverse...
Advanced Search
Add Data
Dataverses (7)
1 to 10 of 16 Results
Sort

McMaster Dataverse **Depositor Checklist**

This checklist augments our more comprehensive McMaster Dataverse Data Deposit Guidelines and McMaster Dataverse Deposit Walkthrough documents. *Items marked with an asterisk * are mandatory for submissions.*

Before Deposit

- ☐ ***README**: Develop a README file to submit alongside dataset [*template, more info*]. You do not need to use the template but information equivalent to sections 1-3 from the template is required.
- ☐ **Documentation**: Gather all the documentation for your dataset, potentially including: data dictionary, data model, code book, interview guide, etc. [*more info*]
- ☐ ***Anonymization/De-Identification:** McMaster Dataverse does NOT accept datasets containing confidential or sensitive information. Remove, replace, or redact data until they are de-identified and non-confidential. [*more info*]
- ☐ ***Ethics Approval**: Confirm you have MREB or HiREB approval to share data (*if applicable*).
- ☐ ***Citation and Credits**: Confirm you have credited, linked, and cited third-party sources, including data, code, or software *(if applicable)*
- ☐ **Dataset organization:** Use consistent file naming and folder organization. A well-structured dataset is easier to understand and share. [*more info*]
- ☐ **Sustainable file formats**: Make your data files accessible long-term and for re-use. [*more info*]

During Deposit

- ☐ ***Account + New Dataset:** Access McMaster Dataverse and set up your account by clicking "Log In" at the top of the page. Click "Add Data" and "New Dataset".
- ☐ ***Data License**: Select an open license under "Dataset Template." [*more info*]
- ☐ ***Add Metadata**: Ensure your data is findable by adding descriptive information.
  - ☐ ***Basic Metadata**: Add Title, Author, Contact, Description, and Subject. [*more info*]
  - ☐ ***Publication:** Enter citation information to Related Publication(s) – *min. title + journal.*
  - ☐ **Time:** Add details to "Time Period Covered" and "Date of Collection" (*if applicable*)
  - ☐ **Place:** Add details to "Geospatial Metadata" (*if applicable*)
- ☐ **Researcher ID**: Connect your research by adding an ORCiD or another researcher ID for yourself and your collaborators. [*more info*]
- ☐ ***Dataset**: Upload data files under the "Files" heading – click "+ Select Files to Add" – *.zip files are unpacked automatically. If you have a lot of files, compress the whole directory in a .zip and upload that file to maintain file/folder structure.*
- ☐ **File Names + Tags**: Modify file names and add text descriptions. If desired, add tags for clarity – select from "Data", "Code," and "Documentation"
- ☐ ***Save Dataset:** Once the dataset is ready, click "Save Dataset"
- ☐ ***Submit for Review:** Submit your dataset for review by RDM Services.

NOTE: If you have chosen not to make data open access, you can restrict access to data files while allowing open access to metadata so your research is still findable. [*more info*]

Depositor "Pre-Flight" Checklist

A joint initiative of McMaster's Faculty of Humanities and McMaster University Libraries

McMaster University

Sherman Centre for Digital Scholarship

Qualitative Data Example

De-identify an interview with a US Senator (public domain)

Quantitative Data Example

Make a README for Nature Vancouver Bird Records (CC BY-NC-SA 4.0)

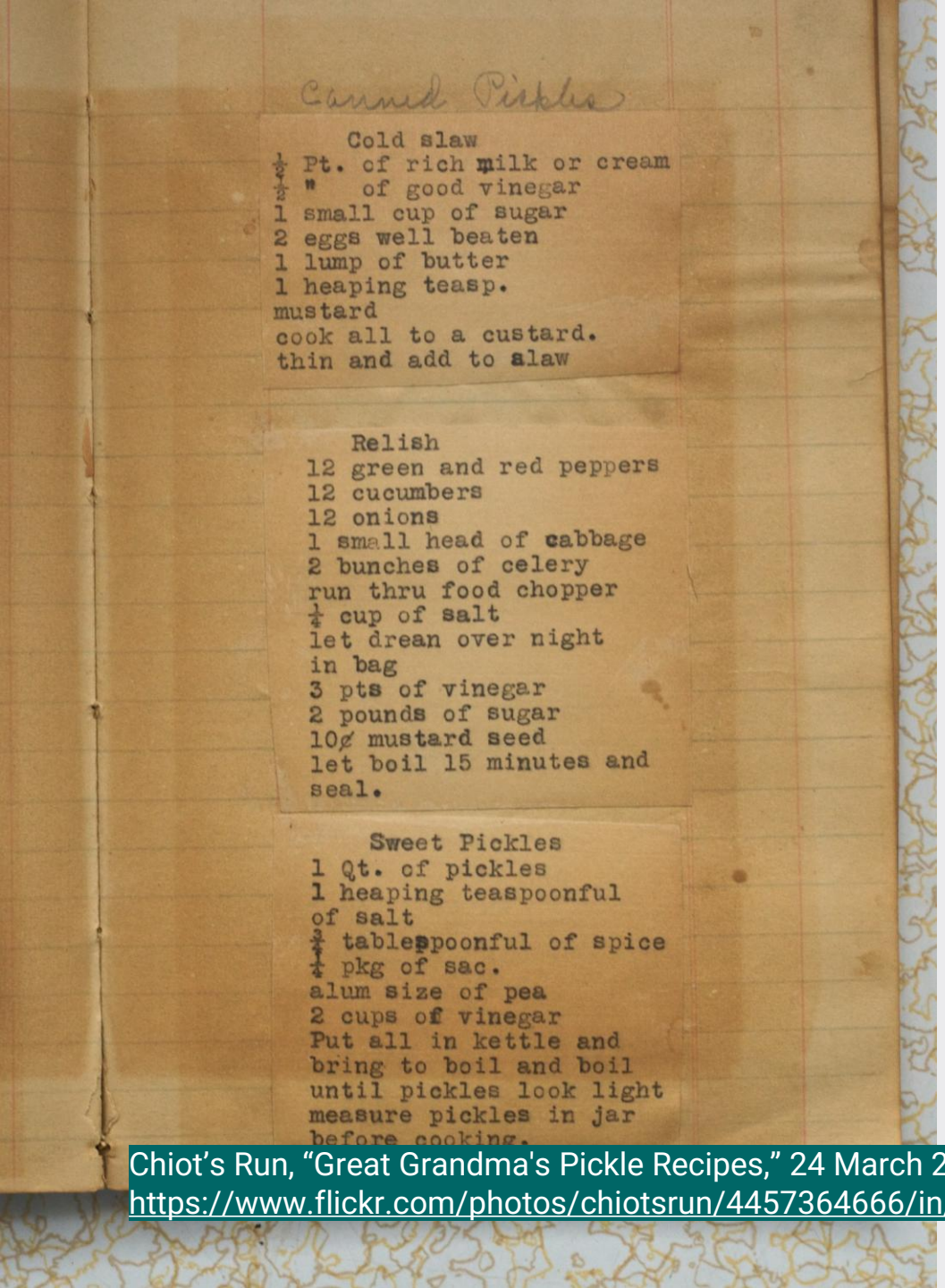
Go here for workshop materials:

https://u.mcmaster.ca/deposit26

Before Deposit

Documentation – README Files, Data Dictionaries, Codebooks, and more.

- ☐ ***README**: Develop a README file to submit alongside dataset [*template*, *more info*]. You do not need to use the template but information equivalent to sections 1-3 from the template is required.
- ☐ **Documentation**: Gather all the documentation for your dataset, potentially including: data dictionary, data model, code book, interview guide, etc. [*more info*]



Documentation: what's in grandma's recipe?

What does looking light mean for pickles? How much is a lump of butter? What is a food chopper and what setting/blade should we use?

- Outline what **variables** mean
- Define **units of measurement**
- Document the **steps** someone would need to take to reproduce your results
- Include a description of any **instruments** and **software**

This example is borrowed from Bibliothèque de l'Université Laval's super cute Gestion des données de recherche series, «Les Métadonnées Et La Documentation», November 3, 2022, https://www.youtube.com/watch?v=0ENz-PIoeOs

Documentation: README

	A	B	C	D	E	F
1	AQ treatment					
2	Experiment 1	mNurr1 (copies)	mBDNF (copies)	mGAPDH (copies)	Normalized mNurr1/mGAPDH	Normalized mBDNF/mGAP
3	Untreated	37350.6000	730.0000	768796.3000	0.1897	0.0010
4	Untreated	37215.9000	755.8000	842354.6000	0.1725	0.0010
5	Untreated	36711.4000	778.1000	763584.7000	0.1878	0.0010
6	10 µM	123274.1000	3420.7000	679878.2000	0.7081	0.0360
7	10 µM	93209.7000	2596.7000	682870.3000	0.5331	0.0270
8	10 µM	109585.1000	2832.3000	752900.1000	0.5684	0.0270
9	Experiment 2	mNurr1 (copies)	mBDNF (copies)	mGAPDH (copies)	Normalized mNurr1/mGAPDH	Normalized mBDNF/mGAP
10	Untreated	54448.4000	846.0000	865315.8000	0.0629	0.0010
11	Untreated	53849.2000	1182.5000	811198.7000	0.0664	0.0010
12	Untreated	53458.0000	1114.0000	774157.1000	0.0691	0.0010
13	10 µM	481003.2000	21085.4000	745384.4000	0.6453	0.0280
14	10 µM	429943.0000	23217.3000			0.0340
15	10 µM	371052.0000	19331.7000			0.0280
16	Depolarization					
17	Experiment 1	mNurr1 (copies)	mBDNF (copies)	mGAPDH (copies)	Normalized mNurr1/mGAPDH	Normalized mBDNF/mGAP
18	Untreated	690325.4167	22004.0010	478178.1033	1.4437	0.0460
19	Untreated	880485.7200	27820.4333	705029.3333	1.2489	0.0395
20	Untreated	701835.1467	24688.7380	391957.8267	1.7906	0.0630
21	Untreated	724400.4200	13203.5198	546267.2667	1.3261	0.0242
22	Depolarized	1538777.3000	34203.4890	434063.7350	3.5450	0.0788
23	Depolarized	1361495.7333	33223.1560	413199.6867	3.2950	0.0804
24	Depolarized	1375530.6667	30856.6790	520936.3667	2.6405	0.0592
25	Depolarized	1526723.9000	31652.7877	507530.3833	3.0081	0.0624
26	Experiment 2	mNurr1 (copies)	mBDNF (copies)	mGAPDH (copies)	Normalized mNurr1/mGAPDH	Normalized mBDNF/mGAP
27	Untreated	556012.8267	14972.0607	730846.4533	1.6990	0.0270
28	Untreated	403071.6006	20023.0030	790077.6667	1.1393	0.0334
29	Depolarized	905870.4300	38844.9600	649662.9167	3.1140	0.0789
30	Depolarized	724244.2867	33138.1200	524733.5000	3.0824	0.0833

raw data

```
This DATSETNAMEreadme.txt file was generated on 2022-03-06 by Margaret Fahnestock



GENERAL INFORMATION

1. Title of Dataset: Nurr1 is not an essential regulator of BDNF in mouse cortical neurons

2. Author Information
        A. Principal Investigator Contact Information
                Name: Margaret Fahnestock
                Institution: McMaster University
                Address: Department of Psychiatry and Behavioral Neurosciences, McMaster University, 1280 Main St. We
                Email: fahnest@mcmaster.ca

        B. Associate or Co-investigator Contac
                Name: Mona Abdollahi
                Institution: McMaster Universi
                Address: Medical Sciences Graduate Program, Faculty of Health Sciences, McMaster University, 1280 Mai
                Email: monaa@mcmaster.ca


3. Date of data collection <2019-09-12 to 2020-11-13>

4. Geographic location of data collection <43.26103813110108, -79.87581398082284, Hamilton, Ontario, Canada>

5. Information about funding sources that supported the collection of the data: Mitacs Research Training Award, Appli


SHARING/ACCESS INFORMATION

1. Licenses/restrictions placed on the data:
```

readme

Documentation

Imagine seeing your data 5 years from now. Can you make sense of what you were doing? Can someone else?

Data documentation supports **interoperability** and **reusability**.

- **README**: A simple text document (.txt) that describes project information, folder hierarchy and file organization, description of important file contents.

https://rdm.mcmaster.ca/sites/default/files/YYYYMMDD_AUTHOR_DATASET_ReadmeTemplate.txt

scds.ca
scds@mcmaster.ca

```
################################################################################
##############
[DATASET TITLE] - This readme file was generated on [YYYY-MM-DD] by
[NAME]
################################################################################
##############
<A readme is a simple text documents that describe the files and
organization of your data, allowing others and your future self to
understand and interpret what you've done. Items that are not relevant
to your dataset may be removed or altered where appropriate.>

<This readme template is provided by McMaster University RDM Services
for projects of any discipline and is adapted from Cornell's readme
template (https://data.research.cornell.edu/content/readme) and
Francesco Varrato, Alain Borel and Chiara Gabella's "README file for
Datasets - Best practices and template"
(https://infoscience.epfl.ch/record/298249)>

<Help text in angle brackets should be deleted before finalizing your
document. [Text in square brackets should be changed for your specific
dataset.]>

-------------------------------

1. GENERAL INFORMATION

Dataset Title:

Overview: <Give a brief overview (2-3 sentences max) to guide the reader
through the available materials and work through your steps from start
to finish.>

<Provide at least two contacts.>
Author/Principal Investigator Information
Name:
ORCID: <Open Researcher and Contributor ID - ORCID is a persistent
identifier to identify authors and contributors and connect your
research outputs, including datasets, into an automatically updated CV
that is unqiue to you apart from any other researchers that may share
your name. Register at https://orcid.org/register.>
Institution:
<ROR: optional - this is a persistent identifier for research
institutions that can be used to connect your research. McMaster's ROR
is https://ror.org/02fa3aq29>
Address: <This should be the institutional address - McMaster's address
```

Documentation

Imagine seeing your data 5 years from now. Can you make sense of what you were doing?

Can someone else?

Data documentation supports **interoperability** and **reusability**.

- **Data Dictionaries**: Document for tabular data describing names, labels, units, and constraints. These help with consistency across multiple researchers; and provide “human-readable” details to support interoperability and reuse.
- **Codebooks:** Like data dictionaries but for survey or statistical data—include layout and structure; codes for questions and answers. REDCap and SPSS automatically generate codebooks.
- *Other documentation could include de-identification protocols, analysis code, interview guides, field notes, and more.*

Documentation – README Files, Data Dictionaries, Codebooks, and more.

- ☑ ***README**: Develop a README file to submit alongside dataset [*template*, *more info*]. You do not need to use the template but information equivalent to sections 1-3 from the template is required.
- ☑ **Documentation**: Gather all the documentation for your dataset, potentially including: data dictionary, data model, code book, interview guide, etc. [*more info*]

File Preparation – Anonymizing, organizing, and more.

- ☐ ***Anonymization/De-Identification:** McMaster Dataverse does NOT accept datasets containing confidential or sensitive information. Remove, replace, or redact data until they are de-identified and non-confidential. [*more info*]
- ☐ ***Ethics Approval**: Confirm you have MREB or HiREB approval to share data (*if applicable*).
- ☐ ***Citation and Credits**: Confirm you have credited, linked, and cited third-party sources, including data, code, or software *(if applicable)*
- ☐ **Dataset organization:** Use consistent file naming and folder organization. A well-structured dataset is easier to understand and share. [*more info*]

File Preparation: Sensitive Data

Sensitive data is any data that would cause harm if released openly. This includes personally identifiable information, personal health information, Indigenous data, ecological or commercial data.

Community Harm: Physical, Psychological, Social, Economic, Legal, Relational.

You cannot share sensitive data openly. If you want to publish or share sensitive data, you have two main options:

1. **Anonymize the dataset:** remove, replace, or redact all sensitive information from datasets prior to upload.
2. Deposit data on a restricted access platform with data sharing agreements.

File Preparation: De-Identification

- Borealis Dataverse does **NOT** accept sensitive data. Remove, replace, or redact data until they are de-identified and non-confidential.
- **Direct Identifiers:** Study Participants are in immediate risk of being re-identified. *Ex. names and initials, relatives and household members, addresses inc. postal codes or zip codes, SIN numbers, emails, photo/audio/video, biometric and genomic data, license numbers, telephone number.*
- **Indirect Identifiers:** Also known as quasi-identifiers. Alone, they can't identify the person, but in combination they can lead to identification. *Gender identity, place of birth, immigration status, household and family composition, role or profession, immigration status, membership in organization.*
- Portage De-Identification Guidance: https://zenodo.org/record/4270551
- Sensitive Data: https://scds.github.io/intro-rdm/sensitive.html

Generation Bass, "Guess-Who-Closeup.jpg" May 4, 2011, https://www.flickr.com/photos/generationbass/5686663807

File Preparation: Ethics Approval

- Informed consent for data sharing - What data may be used for – **defined, extended, broad use**?
- Ideally – get permission before data collection.
- What if I didn't get consent?
  - Consult with MREB or HIREB!
  - TCPS2 Guidance on Depositing Existing Data in Public Repositories: https://ethics.gc.ca/eng/depositing_depots.html

Research Data Management Language for Informed Consent

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

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

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

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

Anonymity vs. Confidentiality

Only research that does not involve the collection of any direct and/or indirect identifiers that can be reasonably traced to individual respondents can be called

guage

in this section to suit the specific needs of their research.

- This study involves collection of anonymous data. This means that no personally identifying information will be collected, including your name or contact information, and that you cannot be identified from the data.

PORTAGE NETWORK / CANADIAN ASSOCIATION OF RESEARCH LIBRARIES 3

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

scds.ca
scds@mcmaster.ca

File Preparation: File and Folder Organization

A good file name makes it easy to find data and keep track of versions. A good file organization system should be **short**, **descriptive**, **standardized**, and **implemented consistently**.

- **File Naming:** Include date, project name, short description, initials of researcher, version number, and other metadata (like location).

  e.g. 2026_10_01_LakeMercury_TestData1_TM_v3.csv

- **Update your README:** Once you've established your organization and naming system, include it in your README file for your research project and make sure it is included prior to data submission.

***Date**: 2026_10_01 (collection date)*
***Project Name**: LakeMercury*
***Short Description**: TestData1*
***Name**: TM (Tracy MacDern)*
***Version Number:** v3*

*Avoid special characters such as & , * % # * ( ) ! @$ ^ ~ ' { } [ ] ?*

File Preparation – Anonymizing, organizing, and more.

- ☑ ***Anonymization/De-Identification:** McMaster Dataverse does NOT accept datasets containing confidential or sensitive information. Remove, replace, or redact data until they are de-identified and non-confidential. [*more info*]
- ☑ ***Ethics Approval**: Confirm you have MREB or HiREB approval to share data (*if applicable*).
- ☑ ***Citation and Credits**: Confirm you have credited, linked, and cited third-party sources, including data, code, or software *(if applicable)*
- ☑ **Dataset organization:** Use consistent file naming and folder organization. A well-structured dataset is easier to understand and share. [*more info*]

Archiving – Choosing what to keep and what formats to keep it in.

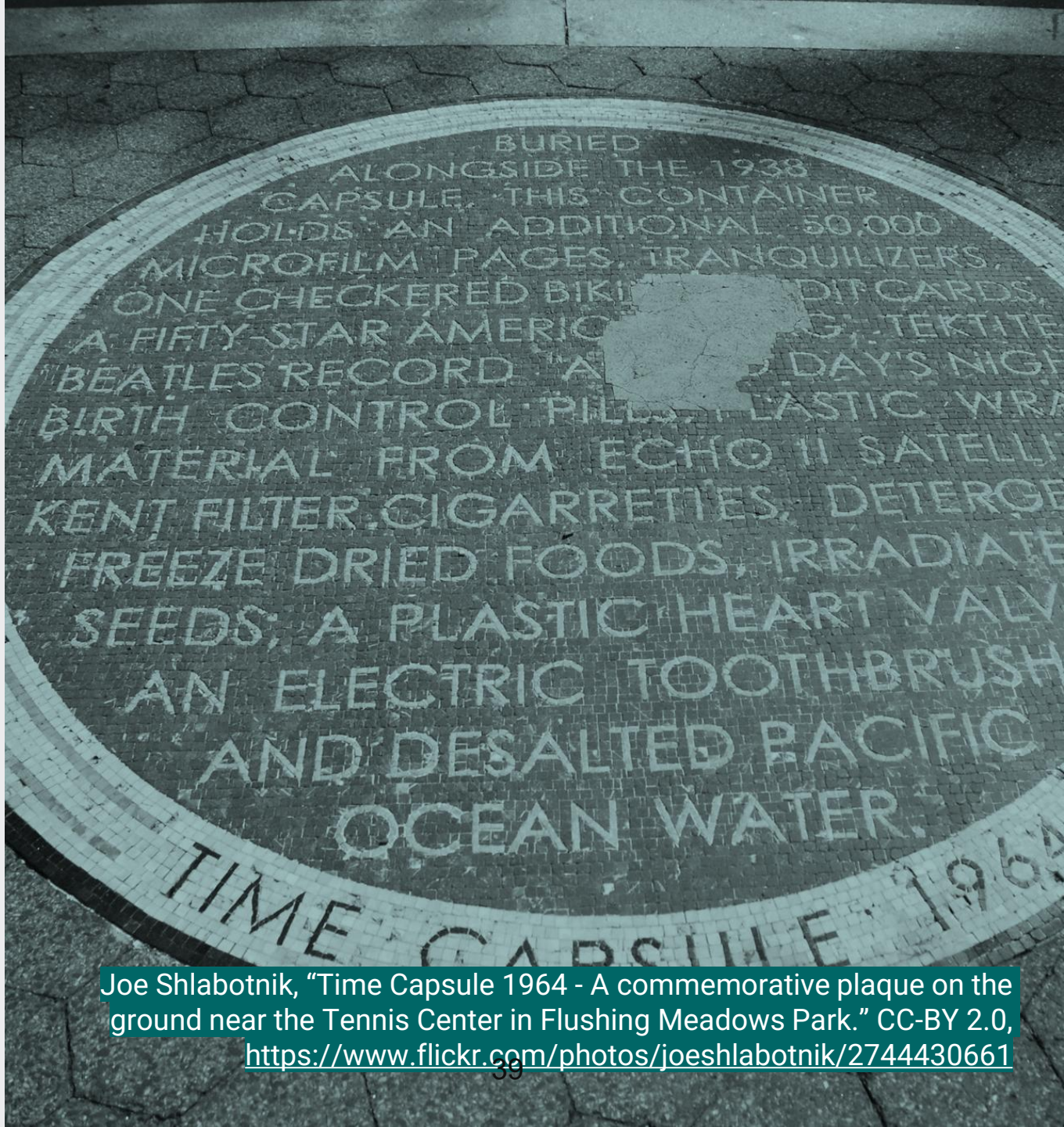
☐ **Sustainable file formats**: Make your data files accessible long-term and for re-use. [*more info*]

Archiving: Appraisal

What are some of the limitations to beaming the entire contents of your GitHub to a generalist repository? Has anyone had their parents give them the entire contents of their childhood bedrooms…just in case you need it?

- **Could/Reuse:** does your dataset support reproducibility? Do you want to reuse it again in the future? Could it be used for teaching and learning?
- **Must/Policy**: Is there a publisher or funder policy asking you to save it?
- **Should/Value:** Is it good and described enough to share? Does it contribute to collective knowledge? Historic significance?
- **Cost**: Are there any costs associated with long-term storage?

Checklist adapted from Angus Whyte, “Five steps to decide what data to keep,” Digital Curation Centre, 31 October 2014, https://www.dcc.ac.uk/guidance/how-guides/five-steps-decide-what-data-keep#7



Joe Shlabotnik, “Time Capsule 1964 - A commemorative plaque on the ground near the Tennis Center in Flushing Meadows Park.” CC-BY 2.0, https://www.flickr.com/photos/joeshlabotnik/2744430661

Archiving: Sustainable File Formats

Other researchers may not have access to any proprietary software you use, so data and metadata should ideally be stored in **sustainable formats**. Look for formats that are:

- Standardized
- Well documented
- In common usage
- Uncompressed

Research instrument files may be manufacturer specific and should be converted to a sustainable format when possible.

See https://site.uit.no/dataverseno/deposit/prepare/#what-are-preferred-file-formats

Examples:

Text: PDF/A, RTF, TXT, XML, LaTeX, md

E-Books: EPUB, PDF/A

Audio: FLAC

Image: TIFF, SVG, PNG

Medical Images: DICOM

Spreadsheet: CSV, TAB

Video: MP4

scds.ca
scds@mcmaster.ca

A joint initiative of McMaster's Faculty of Humanities and McMaster University Libraries

McMaster University

Sherman Centre for Digital Scholarship

Archiving – Choosing what to keep and what formats to keep it in.

☑ **Sustainable file formats**: Make your data files accessible long-term and for re-use. [*more info*]

During Deposit

- [ ] ***Account + New Dataset:** Access McMaster Dataverse and set up your account by clicking "Log In" at the top of the page. Click "Add Data" and "New Dataset".

*Account + New Dataset: Access McMaster Dataverse and set up your account by clicking "Log In" at the top of the page. Click "Add Data" and "New Dataset".

Read full Description [+]

Search this dataverse...

Advanced Search

Add Data

New Dataset

Sort

Dataverses (7)

Datasets (13)

Files (15)

Dataverse Category

Research Group (3)

Laboratory (2)

Research Project (1)

Publication Year

2024 (1)

2022 (2)

2021 (1)

1 to 10 of 20 Results

Replication Data for:

Mar 7, 2024

Pratt, Isaac, 2024, "Replication Data for:", https://doi.org/10.80240/FK2/9MI1DN, Borealis, V1, UNF:6:rmXNGPcxsgkbb3KcwSFVWA== [fileUNF]

test

HOLD Draft Unpublished

Feb 26, 2024

Evering, Danica, 2024, "HOLD", https://doi.org/10.80240/FK2/GA2K9W, Borealis, DRAFT VERSION

Findable + Reusable – Licenses, metadata, and keywords.

- ☐ ***Data License**: Select an open license under "Dataset Template." [*more info*]
- ☐ ***Add Metadata**: Ensure your data is findable by adding descriptive information.
  - ☐ ***Basic Metadata**: Add Title, Author, Contact, Description, and Subject. [*more info*]
  - ☐ ***Publication**: Enter citation information to Related Publication(s) – *min. title + journal.*
  - ☐ **Time:** Add details to "Time Period Covered" and "Date of Collection" (*if applicable*)
  - ☐ **Place:** Add details to "Geospatial Metadata" (*if applicable*)

Metadata: **A license for my data?**

If you don't have a license for your data or code, it falls under the default copyright laws. This means nobody else can copy, distribute, or modify your work without being at risk of violating your copyright.

Open sharing needs an open license, which come in a few flavors:

- Most open licenses are from **Creative Commons (CC)** - (creativecommons.org)
- **Public Domain** means that you are releasing your data with no restrictions.
- **Attribution** ("CC-BY") licenses add a requirement that anyone using the data gives you credit and link to the original dataset.
- Other clauses include **Non-Commercial** ("NC") and **Share-alike** ("SA") restrictions

Creative Commons, fixed by Quibik.

scds.ca
scds@mcmaster.ca

A joint initiative of McMaster's Faculty of Humanities and McMaster University Libraries
McMaster University
Sherman Centre for Digital Scholarship

Borealis > McMaster University Dataverse >

Host Dataverse

Changing the host dataverse will clear any fields you may have entered data into.

McMaster University Dataverse

Dataset Template

Changing the template will clear any fields you may have entered data into.

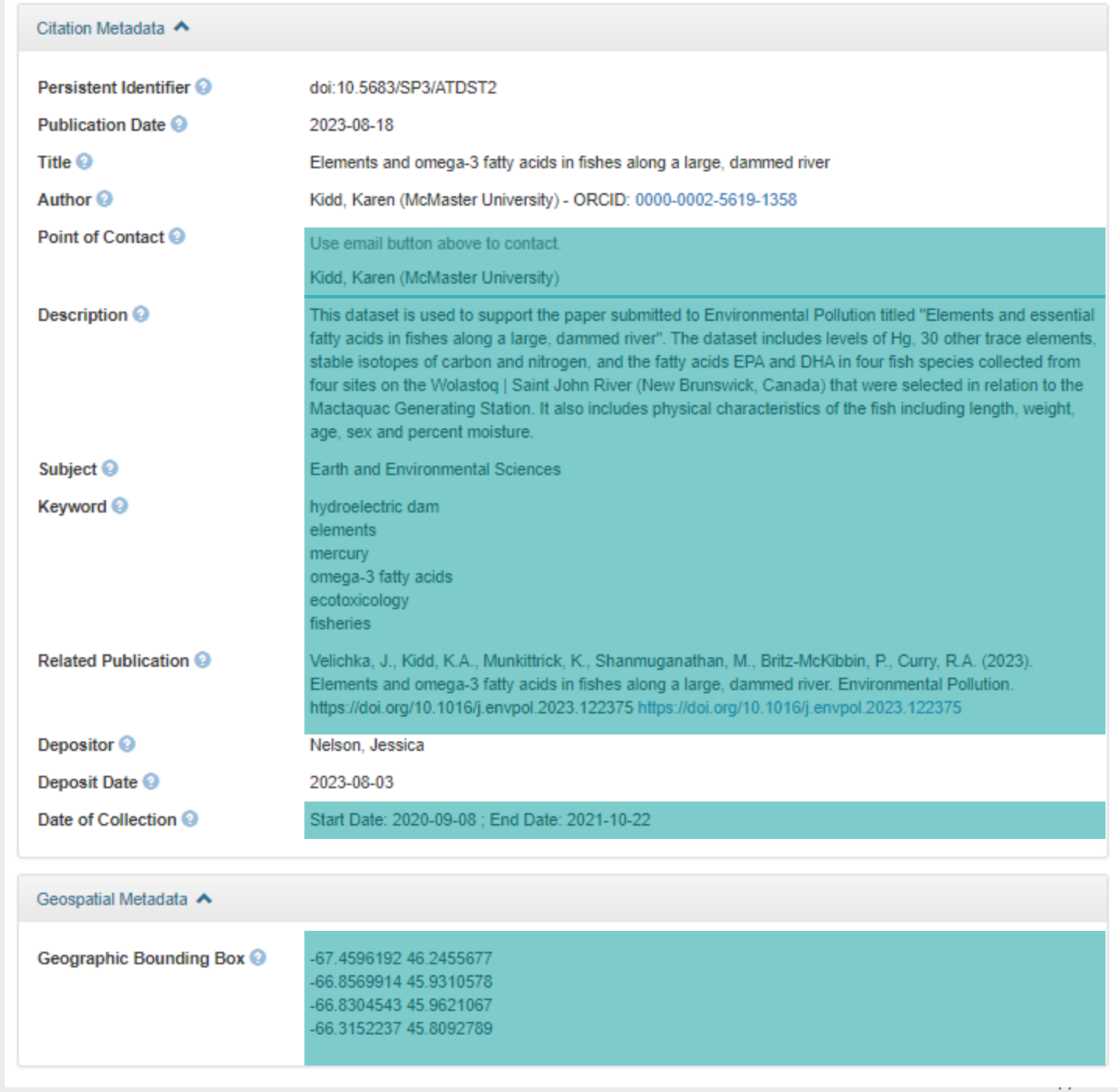
CC0 1.0 Public Domain

*A

- CC0 1.0 Public Domain
- CC BY 4.0 Attribution
- CC BY-NC 4.0 Attribution Non Commercial
- CC BY-NC-ND 4.0 Attribution Non Commercial No derivative/Pas de modification
- CC BY-NC-SA 4.0 Attribution Non Commercial Share Alike/Partage même conditions
- CC BY-SA 4.0 Attribution Share Alike/Partage même conditions
- Custom License Licence personnalisée

Metadata

- Supports Discoverability and Reuse
- Include metadata with your data deposit:
  - Your contact information and affiliation
  - Link to the associated **publication** (if there is one) and its DOI
  - A clear **description** of the data and **keywords**
- Other metadata that might be relevant:
  - **Geospatial** coverage of the data
  - **Time period** covered by the data



Citation Metadata

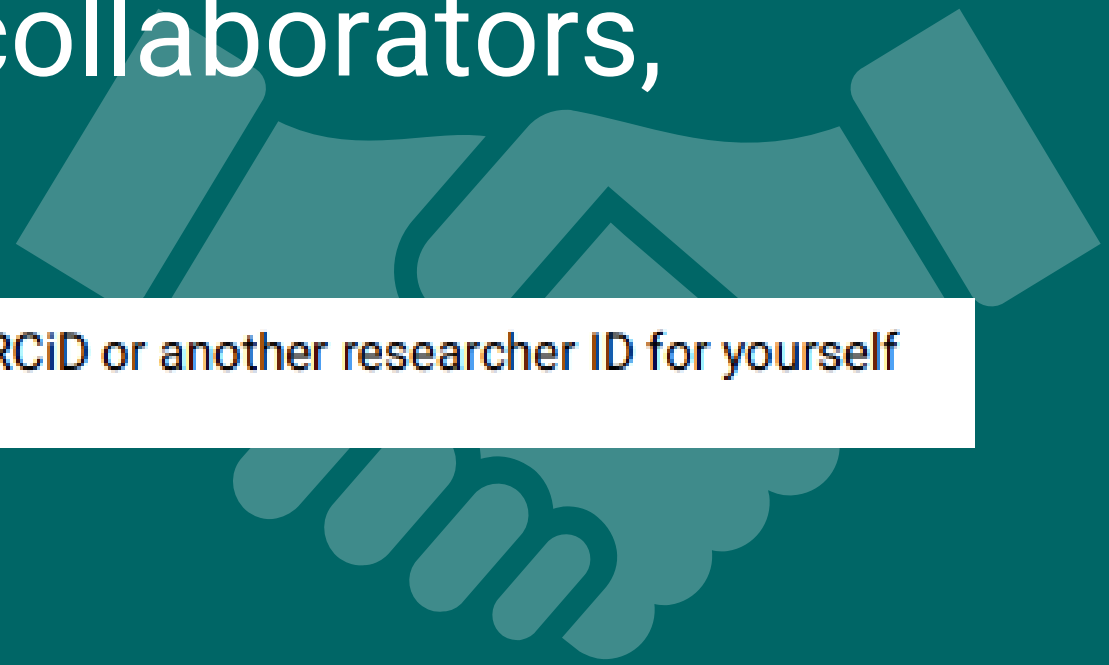
Persistent Identifier	doi:10.5683/SP3/ATDST2
Publication Date	2023-08-18
Title	Elements and omega-3 fatty acids in fishes along a large, dammed river
Author	Kidd, Karen (McMaster University) - ORCID: 0000-0002-5619-1358
Point of Contact	Use email button above to contact. Kidd, Karen (McMaster University)
Description	This dataset is used to support the paper submitted to Environmental Pollution titled "Elements and essential fatty acids in fishes along a large, dammed river". The dataset includes levels of Hg, 30 other trace elements, stable isotopes of carbon and nitrogen, and the fatty acids EPA and DHA in four fish species collected from four sites on the Wolastoq | Saint John River (New Brunswick, Canada) that were selected in relation to the Mactaquac Generating Station. It also includes physical characteristics of the fish including length, weight, age, sex and percent moisture.
Subject	Earth and Environmental Sciences
Keyword	hydroelectric dam elements mercury omega-3 fatty acids ecotoxicology fisheries
Related Publication	Velichka, J., Kidd, K.A., Munkittrick, K., Shanmuganathan, M., Britz-McKibbin, P., Curry, R.A. (2023). Elements and omega-3 fatty acids in fishes along a large, dammed river. Environmental Pollution. https://doi.org/10.1016/j.envpol.2023.122375 https://doi.org/10.1016/j.envpol.2023.122375
Depositor	Nelson, Jessica
Deposit Date	2023-08-03
Date of Collection	Start Date: 2020-09-08 ; End Date: 2021-10-22

Geospatial Metadata

Geographic Bounding Box	-67.4596192 46.2455677 -66.8569914 45.9310578 -66.8304543 45.9621067 -66.3152237 45.8092789

scds.ca
scds@mcmaster.ca

Citation + Credit – Linking and giving credit to data sources, collaborators, and publications.

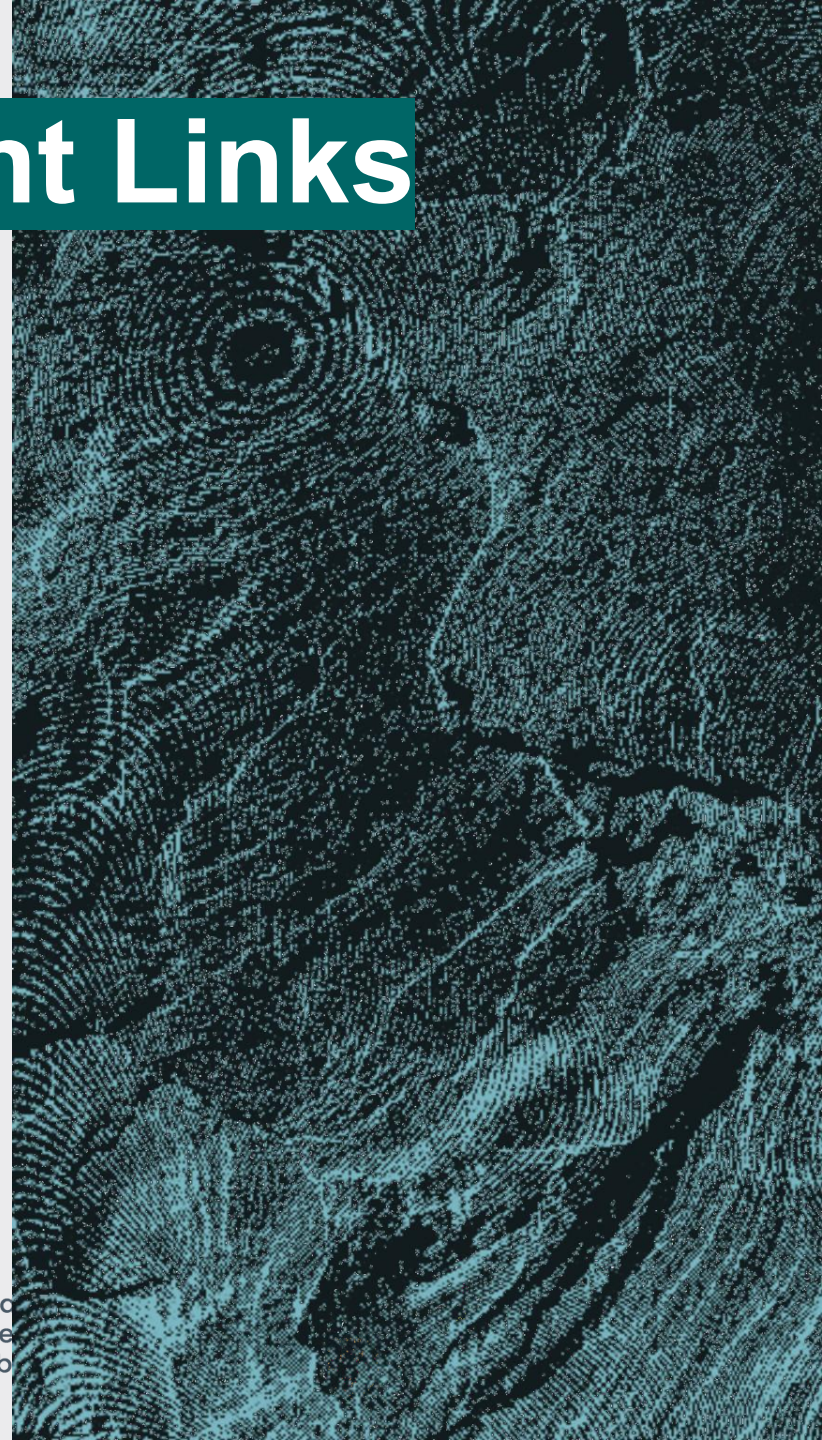


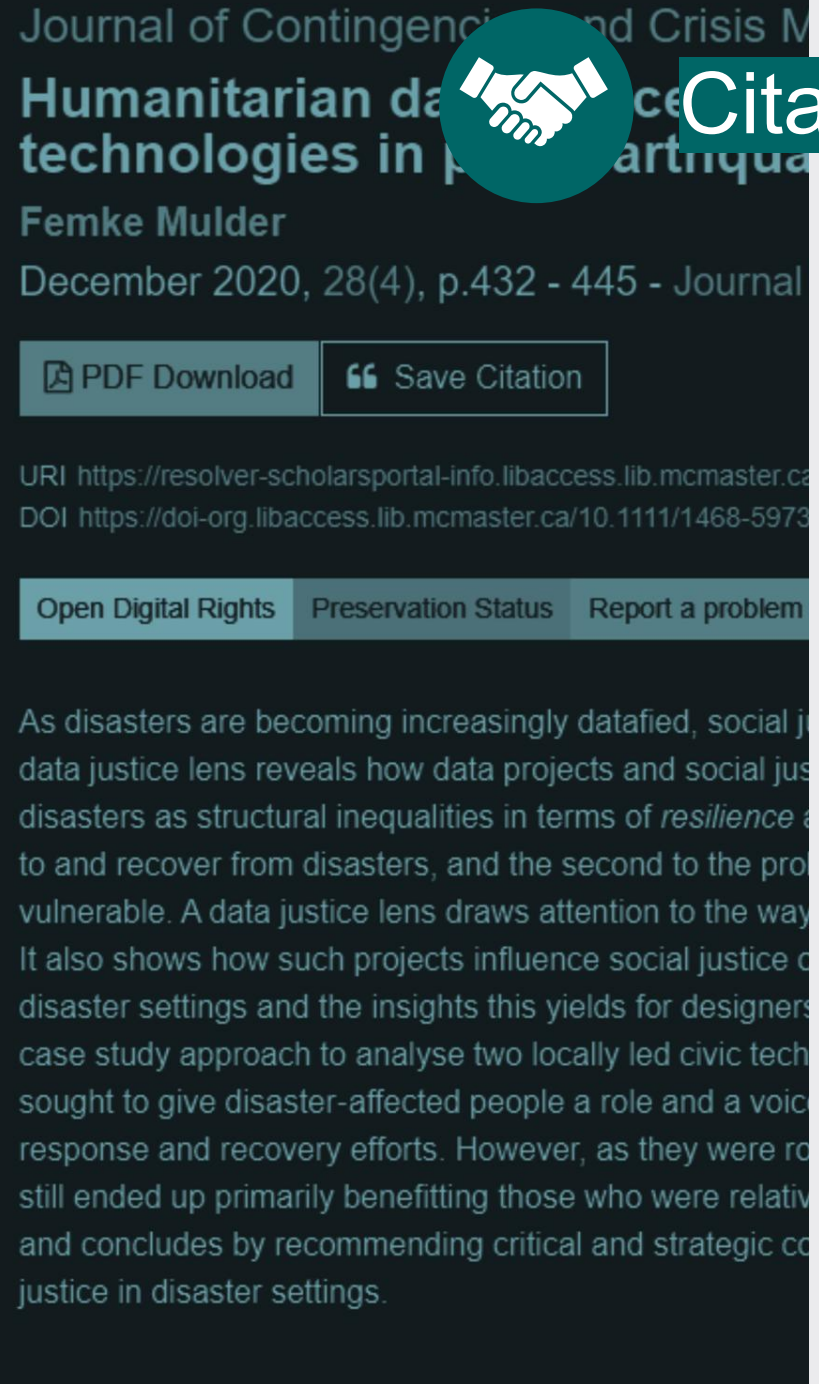
☐ **Researcher ID**: Connect your research by adding an ORCiD or another researcher ID for yourself and your collaborators. [more info]

Citation + Credit: **Permanent Links**

[8] Attendees of the NDSF Summit. (2019). Kanata Declaration. Zenodo. http://doi.org/10.5281/zenodo.3234815

- **Persistent Identifiers (PIDs)** Persistent Identifiers (PIDs) are unique and immovable touchpoints or anchors; unique links that will never expire.
- They help facilitate links between researchers, publications, and more.
- Clearer for **humans** reading them, and make more stable, interconnected systems for **computers** to read.
- Concept has been around forever – ISBNs are a unique identifier!





Citation + Credit: **DOI – Digital Object Identifier**

ex. https://doi.org/10.5683/SP2/VAPPJQ

- **Content**: articles, books, ebooks, datasets, publications, audio/visual content, software
- Not-for-profit organization, registration authority for the ISO standard
- **Crossref** and **DataCite** are foundation members and can create DOIs
- ***Link them!*** *Add DOIs for your publication, software, and any secondary data used to your data submission (and vice versa!)*

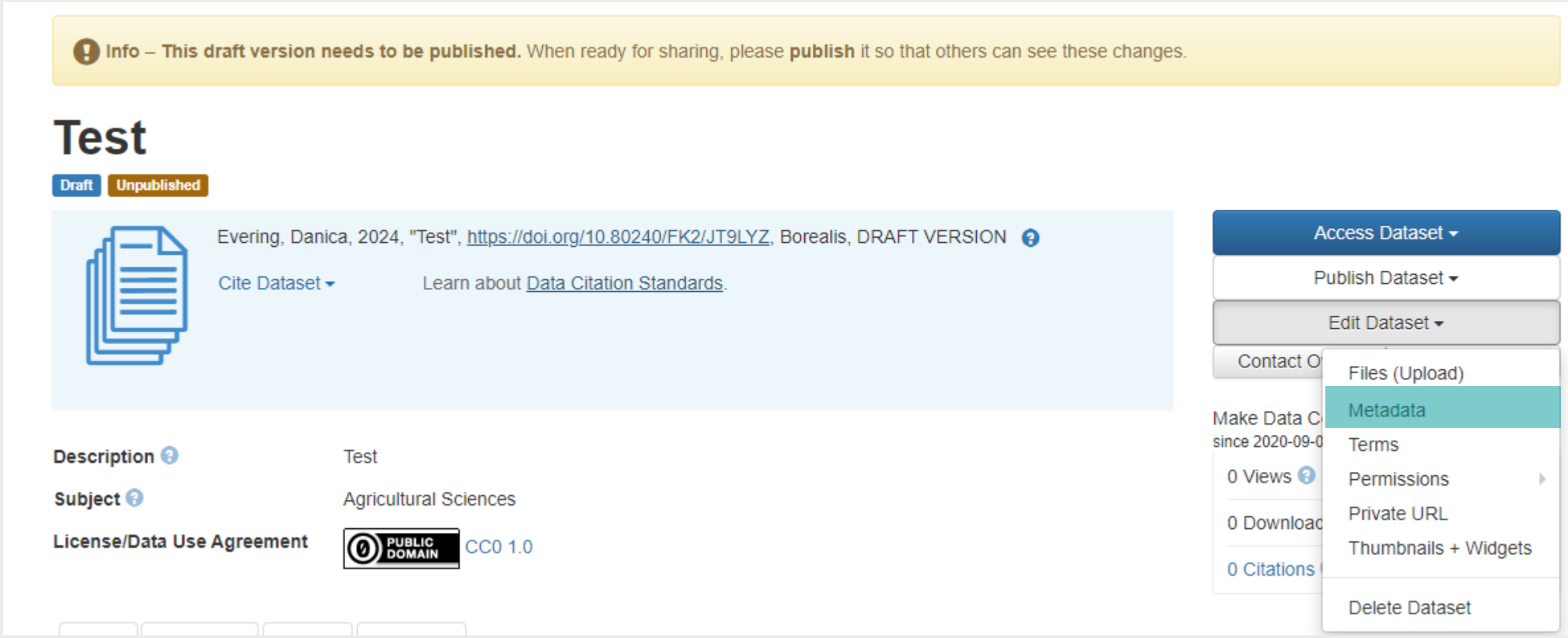
Citation + Credit: **ORCID - Open Researcher + Contributor ID**

ex. https://orcid.org/0000-0001-6326-7659

- **Content**: researchers, authors, contributors
- Open, global, community-built not-for-profit organization.
- Distinguish yourself from scholars with the same name - *Is John Smith @ McMaster = John Smith @ Queen's?*
- Connect your datasets, code, and publications.
- ***Link them!*** *Add ORCiDs for all your team members to your dataset*.

researchers
contributors
authors

Contributor Roles Taxonomy, "CRediT: 14 Contributor Roles," 2023, accessed February 10, 2023, https://credit.niso.org/
Image by Christina @ WOCinTech.com on Unsplash.


Info – This draft version needs to be published. When ready for sharing, please publish it so that others can see these changes.
Test
Draft
Unpublished
Evering, Danica, 2024, "Test", https://doi.org/10.80240/FK2/JT9LYZ, Borealis, DRAFT VERSION
Cite Dataset
Learn about Data Citation Standards.
Description
Test
Subject
Agricultural Sciences
License/Data Use Agreement
PUBLIC DOMAIN
CC0 1.0
Access Dataset
Publish Dataset
Edit Dataset
Contact O
Files (Upload)
Metadata
Terms
Permissions
Private URL
Thumbnails + Widgets
Delete Dataset
Make Data C
since 2020-09-0
0 Views
0 Download
0 Citations

Citation + Credit: **Edit Metadata**

- **Author**: May be different than the paper – who has contributed significantly to the creation of the data?
- **Point of Contact:** Consider adding >2 for longevity
- **Related Publication:** Link publication DOIs
- **Contributor**: Did anyone else have a minor role in the creation of the dataset?
- **Related Material, Related Dataset, Data Source**
- **Other Metadata**: Time Period + Geospatial Metadata

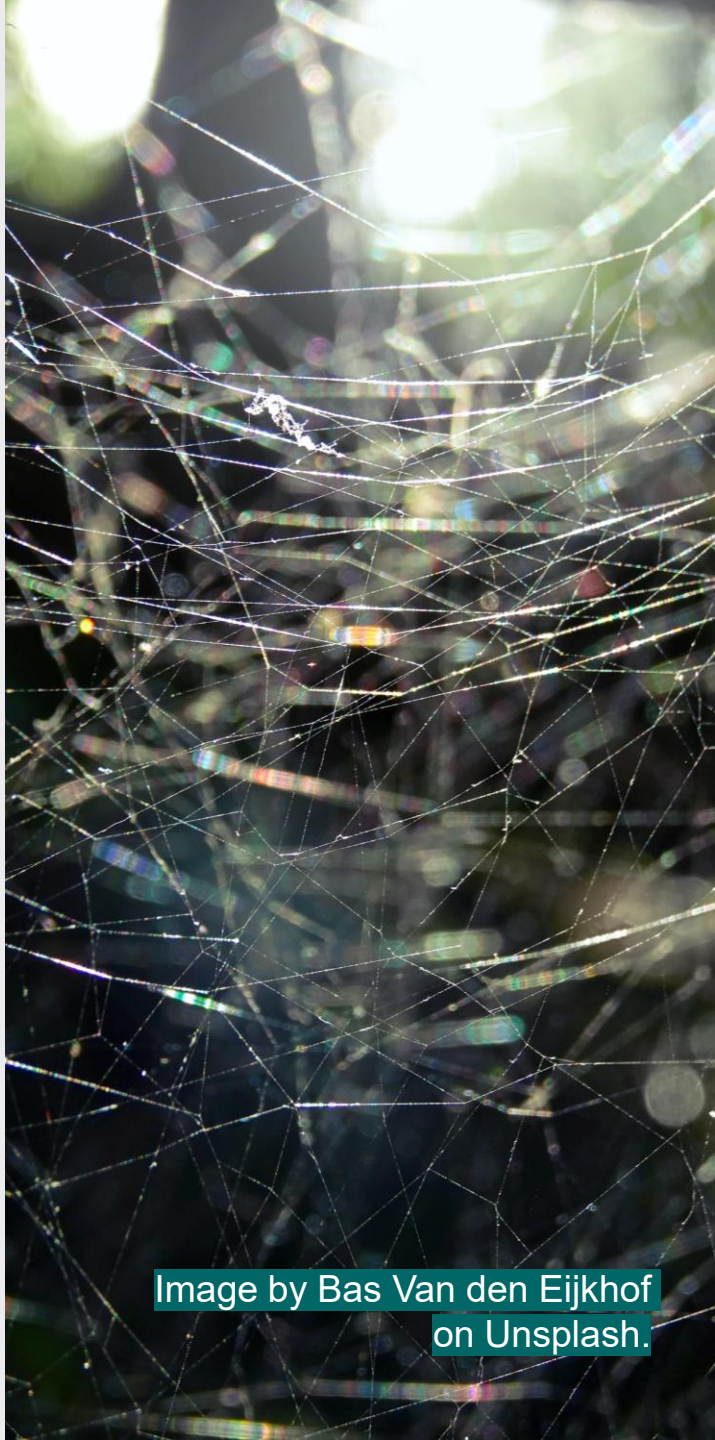


Image by Bas Van den Eijkhof on Unsplash.

1 File

Edit

File Name 2024-06-17_Test-Datafile.xlsx

File Path

MS Excel Spreadsheet
MD5: 0bb8c68ec7f5c9103f2eaa15a0fab3cc

Description Add file description...

Edit Options
Provenance
Tags

Metadata Tip: After adding the dataset, click the Edit Dataset button to add more metadata.

Save Dataset Cancel

Next steps

- When you finish your dataset, save it!
- Click “Submit for Review” to notify RDM Services that your dataset is ready for curation (for us to review it!)
- If you have more questions, reach out to us by email or set up an appointment

Sherman Centre for Digital Scholarship

Research Data Management Links

Send RDM Services an email:

rdm@mcmaster.ca

Review resources on our webpage:

https://rdm.mcmaster.ca

Join our Community of Practice:

https://u.mcmaster.ca/rdm-community

Make an appointment:

https://u.mcmaster.ca/rdm-appointments

SCDS Links

Send SCDS an Email:

scds@mcmaster.ca

Subscribe to our Newsletter:

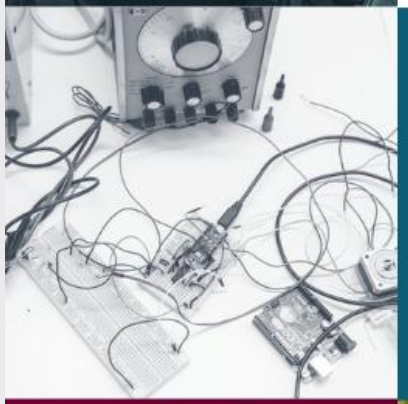
https://u.mcmaster.ca/sign-up

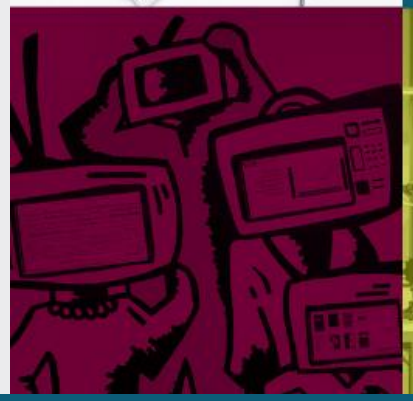
Register for a Workshop:

https://u.mcmaster.ca/scds-workshops

Schedule a Consultation:

https://libcal.mcmaster.ca/appointments





McMaster University Library