

Open Access to Knowledge Part 2: Sharing Your Research Data

Mélanie Brunet
Jeanette Hatherill
Chantal Ripp

University of Ottawa Library



Unless otherwise specified, the contents of this presentation are made available under a [Creative Commons Attribution 4.0 International licence](https://creativecommons.org/licenses/by/4.0/).

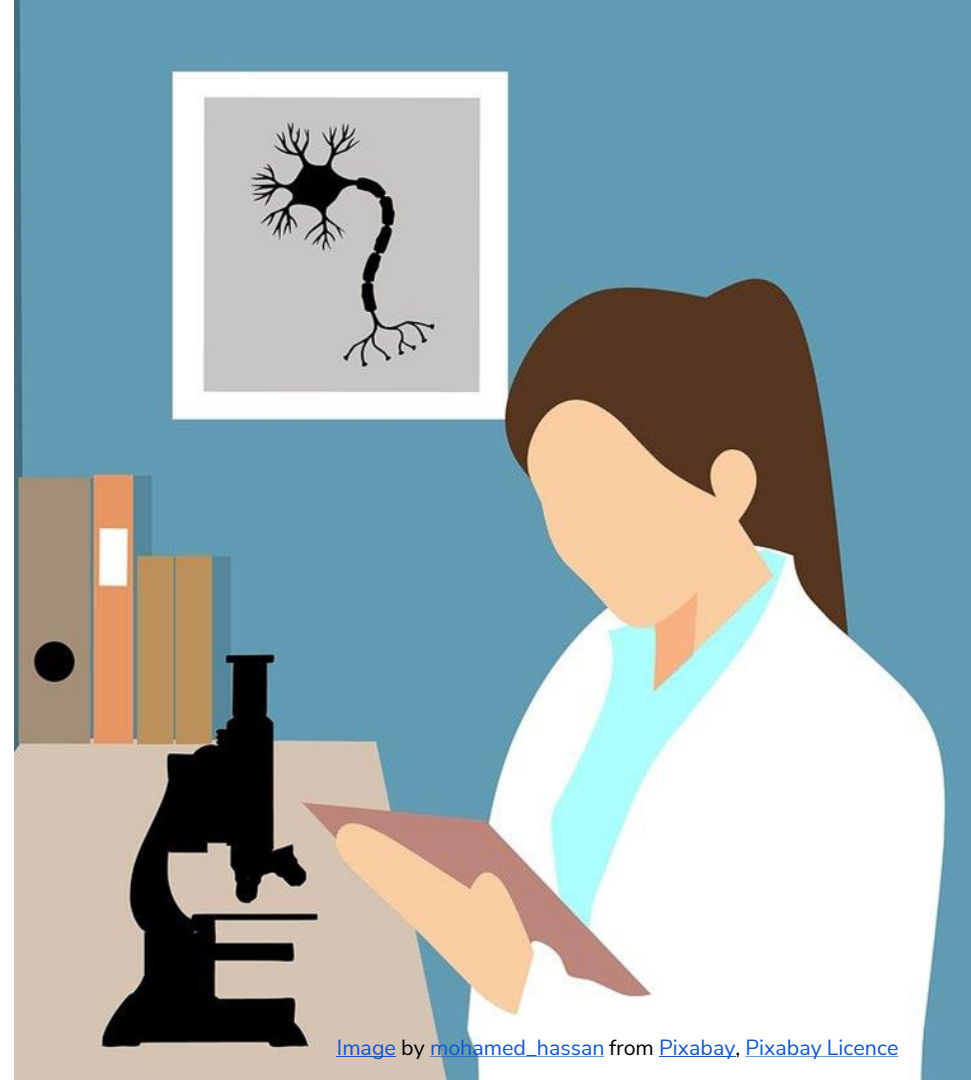
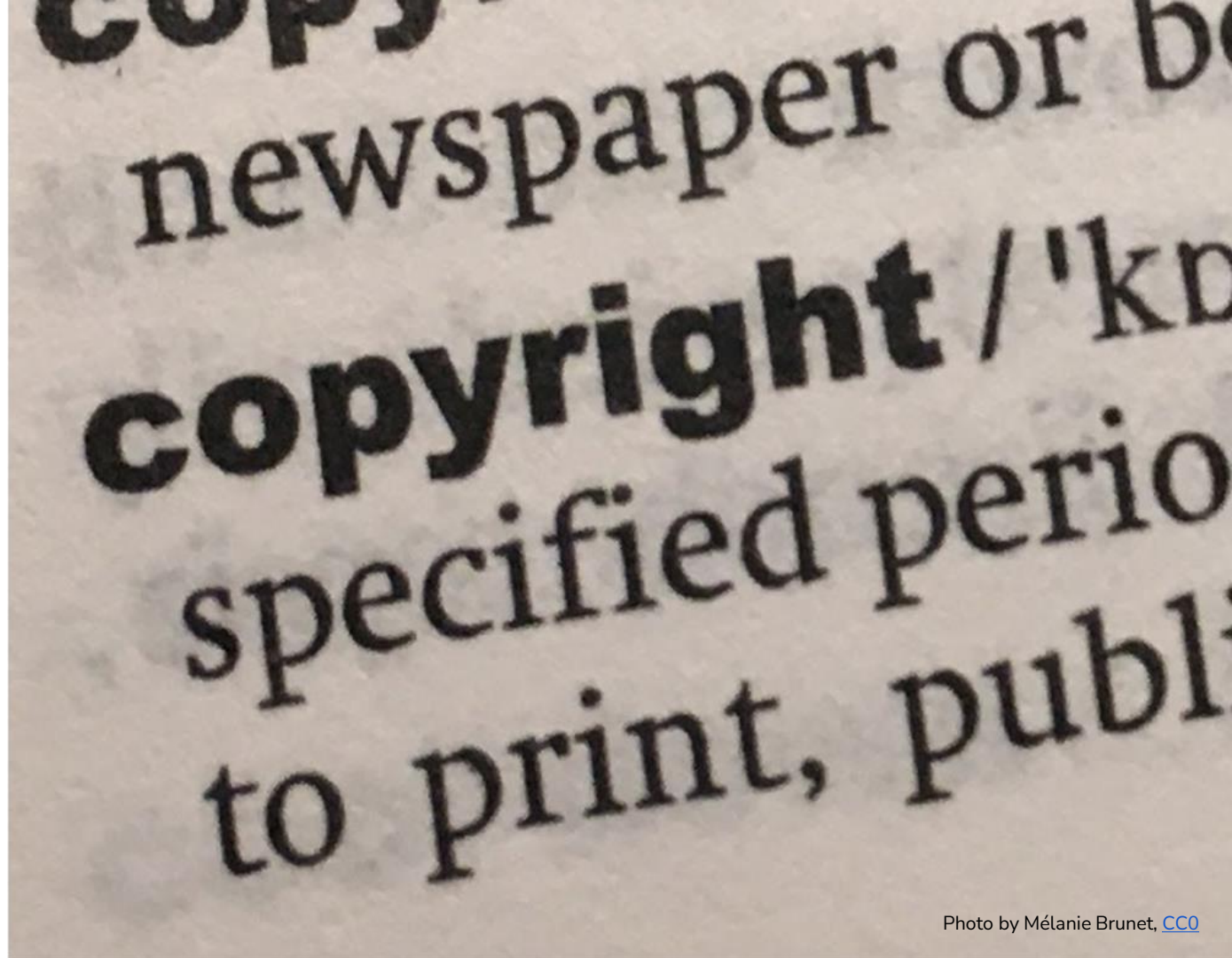


Image by [mohamed_hassan](https://pixabay.com/users/mohamed_hassan/) from [Pixabay](https://pixabay.com/), Pixabay Licence

By the end of this webinar you will:

- Be familiar with copyright and data ownership
- Plan to share your research data
- Make your data and supporting documentation available in a recognized repository
- Understand how to apply open licences to your data

1. Copyright



Copyright Basics



Copyright...

- protects the **original expression** of ideas or facts **fixed** in a tangible medium
- is **automatic** upon fixation

raw/factual data vs. compilation of data

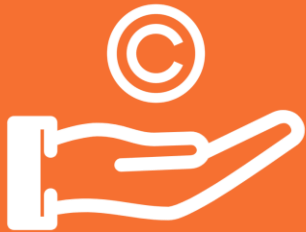
Is data protected by copyright?



It could!

Not protected by ©	Could be protected by ©
Datum/raw data (ex. a number, a measurement)	Data representations (ex. tables and graphs)
	Data sets
	Purchased data (with licence terms)
	Datum that by itself is a literary, musical, dramatic or artistic work

Is there copyright protection and who owns it?



1. Determine if data protected by copyright
2. If yes, who owns it? Don't assume it's you:
 - a. Employment contracts and institutional policies
 - b. Disciplinary conventions and practices in attributing authorship
 - c. Policies of funding agency or organisation
 - d. Licensing conditions of purchased data

➤ Decision to share data and level of openness will depend on answers to these questions.

➤ If there is copyright but you do not own it, data cannot be made more open without permission of copyright owner.

Copyright and sharing data as supplementary materials to a publication



In subscription-based journals...

Supplementary materials require copyright transfer:

- May not have authority to transfer © on data
- Data will remain behind paywall and therefore not available openly

➔ Choose data repositories instead

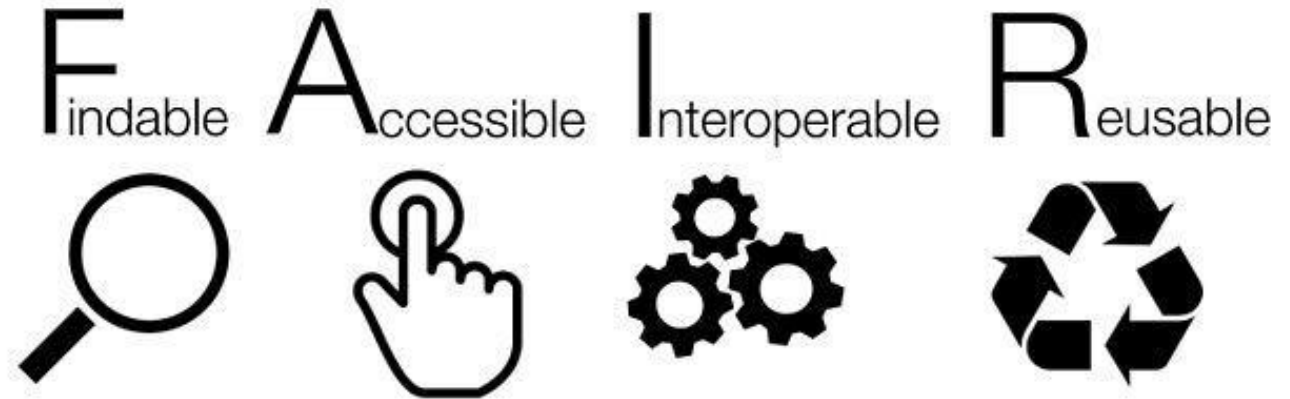
2.

Begin with
the end in
mind



Jones, Sarah, &
Grootveld, Marjan.
(2017, November).

[How FAIR are your
data?](#)



[FAIR data principles](#) by SangyaPundir, Wikimedia Commons, [CC BY-SA 4.0](#)

Plan to share your research data



Data Management Plan (DMP) is a document that assist researchers in determining the costs, benefits and challenges of managing data.



➡ Create at the beginning of a project and update it as project evolves.

DMP Assistant



Use a template - Portage's [DMP Assistant](#)

Project Details Plan overview Write Plan Share Download

expand all | collapse all 1/20

Data Collection (1 / 3)

This section addresses data collection issues such as data types, file formats, naming conventions, and data organisation – factors that will improve the usability of your data and contribute to the success of your project.

What types of data will you collect, create, link to, acquire and/or record?

B *I* [List Icons] [Link Icon] [Table Icon]

Three types of data will be collected -- survey data, interview data, and documents. The initial study phase will collect quantitative survey data based on the Physician Orientation Profile (POP) questionnaire, a tool for measuring how physician-researchers approach their patients and the clinical trial they are participating in as investigators (Taylor & Kelnner, 1987). The adapted POP questionnaire includes two parts. The first collects demographic information and information about professional activities in both closed question and open text formats. The second part includes 45 binary response questions aligned to a therapist-experimenter continuum.

Qualitative data will be collected using a semi-structured interview protocol. Audio recordings will be taken of the interviews. The interviews will be directly transcribed and the interview scripts and analytic memos will be added to the database. Code books from the thematic analysis of the interviews will be created and added to the study database. Additionally, if follow-up interviews are performed, they will be similarly transcribed and added to the database. Publicly available documents will be collected to triangulate the data in the investigator cases. Data will be gathered from ClinicalTrials.gov, professional society websites (if publicly available without membership), published trial reports from PubMed, and institutional/practice websites. Code books from the content analysis of the collected documents will be created and added to the study database.

Guidance Comments

Portage

Examples: numeric, images, audio, video, text, tabular data, modeling data, spatial data, instrumentation data.

DMP Examples



Look at other DMPs as models

» [Portage Network Exemplar DMPs](#)

» [Digital Curation Centre - Example DMPs and Guidance](#)

Storage and Backup

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

All data storage and back-up procedures will be clearly outlined within the project's data collection policies and procedures which will be developed prior to data collection.

Survey data will be collected using REDCap, a secure data collection and management software hosted by the Women & Children's Health Research Institute (WCHRI) at the University of Alberta. REDCap servers undergo regular (daily, weekly, monthly) backups. Secure transferring of data from the REDCap platform to our virtual cloud-based research environment will occur using both encryption and a secure File Transfer Protocol platform. Upon being exported from REDCap, survey data will be immediately ingested into our research project space located on Compute Canada's cloud platform and which has regularly scheduled backup processes in place.

Qualitative interviews will be conducted using encrypted digital voice recorders. Upon completion of interviews they will be securely transferred within 48 hours to the virtual research project space located on Compute Canada's cloud platform. Once the audio interviews are uploaded to the cloud platform, they will be deleted from the digital voice recorders.

How will you ensure that sensitive data is stored securely and only accessible to the research team during the research project?

All data will be securely stored in our virtual project space located on Compute Canada's cloud platform and these will be accessible only by approved researchers, trainees, and study staff. Access to the platform is securely password protected, with access rights ultimately approved by the Principal Investigator and assigned by Compute Canada.

For more information regarding this service provided by Compute Canada see: <https://www.computeCanada.ca/research-portal/accessing-resources/rapid-access-service/>

Doiron, James. (2020, September 8). Data Management Plan Exemplar #3: Mixed Methods Fictional Exemplar. Zenodo. <http://doi.org/10.5281/zenodo.4019563>

Sensitive Data Toolkit for Researchers



Access to individual participant data and trial documents should be *as open as possible and as closed as necessary*

Understand the risk level for human participant research data to make decisions with respect to its management, deposit, and appropriate access/future use.

Toolkit:

- Part 1: a [Glossary of Terms for Sensitive Data used for Research Purposes](#)
- Part 2: a [Human Participant Research Data Risk Matrix](#)
- Part 3: and a [document providing samples of Research Data Management Language for Informed Consent](#).

Anonymize personal information



Minimize disclosure risk when sharing data collected from human participants by applying de-identification practices

➤ Consult Portage's [De-identification Guidance](#)

Fee-based services for de-identification

- [d-wise](#) (American & European offices)
 - Offering free [anonymization services to researchers](#) who deposit individual participant-level data from COVID-19 clinical trials in Vivli.
- Inter-university Consortium for Political and Social Research(ICPSR)(Archive headquartered at University of Michigan
 - Purchase the Professional Curation package, see [Deposit Data with ICPSR](#)
- [Privacy Analytics](#) (Ottawa-based company)
 - Preview datasets as part of their Data Privacy Validation Services.

File formats



When choosing a format for your data you need to consider later lifecycle activities

Proprietary Format	Preferred Format
Excel (.xls, .xlsx), SPSS	Comma Separated Values (.csv or .tsv) ASCII
Word (.doc, .docx)	plain text (.txt), XML or if formatting is needed, PDF/A (.pdf)
PowerPoint (.ppt, .pptx)	PDF/A (.pdf)
Photoshop (.psd)	TIFF (.tif, .tiff)
Quicktime (.mov)	MPEG-4 (.mp4)

» Not sure about a file format? Check OpenAIRE's [Data formats for preservation guide](#) and Library and Archives Canada's [Advice on the digital file formats to be used when transferring information](#)

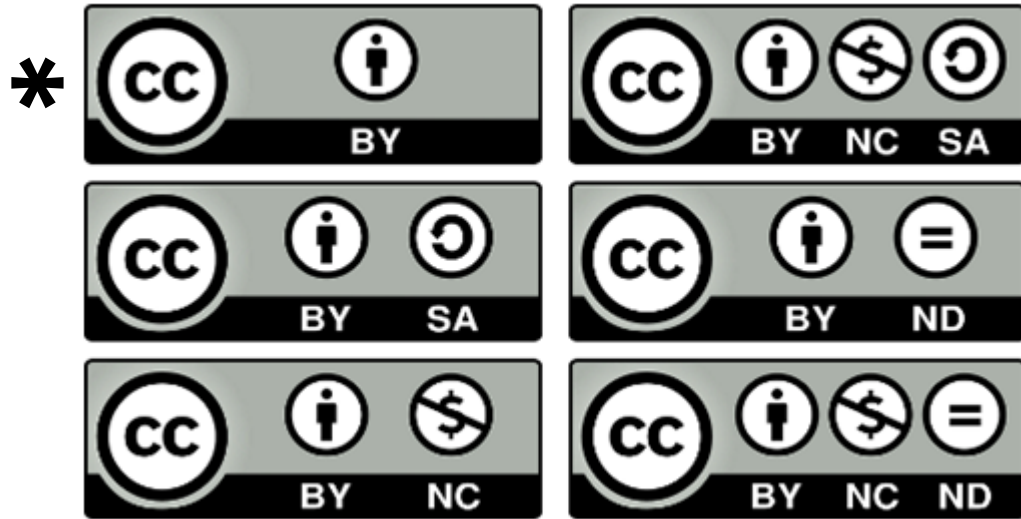
3.

Open
Licences



What is a licence?

- “A licence allows someone else to use a work for certain purposes and under certain conditions.
- The copyright owner still retains ownership.”

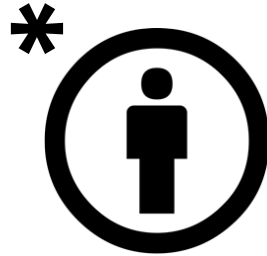


Creative Commons Attribution 4.0 International Licence

CC BY 4.0

About the licences: <https://creativecommons.org/licenses/?lang=en>

Creative Commons licences conditions



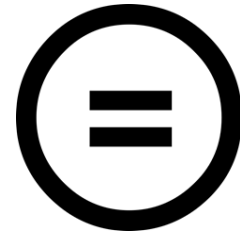
Attribution (BY)



ShareAlike (SA)



NonCommercial (NC)



NoDerivatives (ND)

CC0 - Public Domain Dedication



CC0

Instrument used by creators to waive their rights and dedicate their work to the public domain right away

More about CC0:

<https://creativecommons.org/share-your-work/public-domain/cc0/>

Software licences



[MIT License](#)



[GNU General Public License 3.0](#)



[Apache License 2.0](#)

[MIT logo](#), Wikimedia Commons, public domain, [GPL logo](#), Wikimedia Commons, public domain, [Apache logo](#), Wikimedia Commons, [Apache License 2.0](#)

Licence comparison chart

Licence	Distribution	Modification	Sublicensing	Attribution
CC BY	Allowed	Allowed	Allowed	Required
CC0	Allowed	Allowed	Not allowed	Not required
MIT	Allowed	Allowed	Allowed	Required
GNU GPL	Allowed	Allowed	Allowed	Required
Apache	Allowed	Allowed	Allowed	Required

- All 5 licences allow for commercial use
- Sublicensing = derivatives can be shared under different licence

Table licensed [CC BY-SA 4.0](#), based on [Comparison of free and open-source software licences](#), Wikipedia, [CC BY-SA 3.0](#)

4.

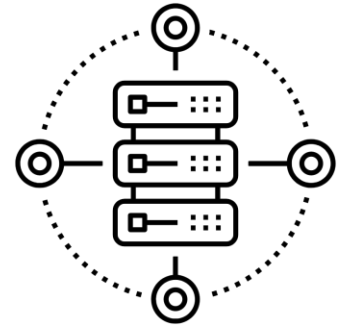
4 Easy Steps to Share Research Data



Why a data repository?

Repositories provide :

- Technical infrastructure
- Visibility of your research data (collected in discovery platforms such as [OpenAIRE COVID-19 Guideway](#))
- Permanent identifier (DOI) and citation of your data
- Access controls



Step 1: Identify a disciplinary or generalist repository



Look for:

- Recommendations or requirements in funder guidelines (e.g. [CIHR Open Access Policy: Annex](#))
- Recommended repositories maintained by journals/publishers (e.g. [Scientific Data](#) and [PLOS ONE](#))
- [FAIRsharing](#) or [re3data.org](#) to consult a list of certified data repositories
- Consult library's [RDM website](#) for supported repositories, including uOttawa Dataverse and Federated Research Data Repository (FRDR)

Disciplinary repositories examples



GenBank

[GenBank](#)



[NDEX v2.5.0](#)



[ImmPort](#)



OPENICPSR



[COVID-19 Data Repository OpenICPSR](#)

Generalist repositories examples



[Scholars Portal Dataverse](#)



[Open Science Framework](#)



[Federated Research Data Repository \(FRDR\)](#)



[Zenodo](#)



[Figshare](#)

Step 2: Deposit dataset(s) and documentati on



Data file(s)

Provide documentation and supporting information (e.g. codebook and/or Readme file; methodologies; data provenance).

- Portage's [Documentation and Supporting Material Required for Deposit](#)

Scholars Portal Database > York University Database > Aaida Mamuji Database > COVID-19 Destigmatizing Chinese Communities Project >

COVID-19 Destigmatizing Chinese Communities Project: Greater Toronto Area Interview Data

Version 1.0

Mamuji, Aaida, Rozdilsky, Jack L., Lee, Charlotte T., Mwarumba, Ngaki, Tubala, Martin, Chu, Temi. 2020. "COVID-19 Destigmatizing Chinese Communities Project: Greater Toronto Area Interview Data", <https://doi.org/10.5663/SP2/GC302H>, Scholars Portal Database, V1

[Cite Dataset](#) - [Learn about Data Citation Standards](#).

Access Dataset ▾

Contact Owner Share

Dataset Metrics ⓘ

13 Downloads ⓘ

Files Metadata Terms Versions

Search this dataset.

Filter by
File Type: All - Access: All -

Sort ▾

1 to 10 of 86 Files		Download
	0_Toronto Key Informant Interview Guide.pdf Adobe PDF - 55.4 KB - Sep 22, 2020 - 9 Downloads MDS: d0d8f6e3951e895116cc2c552c7e01	
	Interview 01 - Public Transcript.rtf Rich Text Format - 1.1 MB - Sep 22, 2020 - 2 Downloads MDS: 067e8e729c308d939a10ebba9a6d	
	Interview 02 - Public Transcript.rtf Rich Text Format - 1.2 MB - Sep 22, 2020 - 0 Downloads MDS: 3717c79e64cf727aa6f7683a821a30	
	Interview 03 - Public Transcript.rtf Rich Text Format - 1.1 MB - Sep 22, 2020 - 0 Downloads MDS: 6ad5a2641cdac2c79929a9697f7e5c	
	Interview 04 - Public Transcript.rtf Rich Text Format - 1.2 MB - Sep 22, 2020 - 0 Downloads MDS: 241923d7406daf8140ec08be6b29a30	
	Interview 05 - Public Transcript.rtf Rich Text Format - 1.1 MB - Sep 22, 2020 - 0 Downloads MDS: 87569391a246e91923c3962e1a851	

Step 3: Add descriptive metadata

Make your research outputs richer, connected and reusable with robust metadata

Recommended metadata fields:

Dataset Name	Author(s)*	Description
Keywords	Date Published	Date Modified
Licence	Funder and Grant #	DOI
Related identifiers (e.g. to publication or code if in another repository like Github)		

*ORCID iD

Step 4: Apply Open Licences to Your Data

Ex.
Scholars
Portal
Dataverse


Files Metadata Terms Versions

Edit Terms Requirements

Terms of Use ^

Waiver ?

Our [Community Norms](#) as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the Dataverse.

Default is CC0 → CC0 - "Public Domain Dedication" 

Want to apply
an open
licence? Click
"Edit Terms
Requirements"

Step 4: Apply Open Licences to Your Data

Ex.
Scholars
Portal
Dataverse

Files Metadata Terms Versions

Edit Terms Requirements

Terms of Use ^

1. Click "No, do not apply CC0"

Waiver ?

Terms of Use ^

Datasets will default to a [CC0 public domain dedication](#). CC0 facilitates reuse and extensibility of research data. Our [Community Norms](#) as well as good scientific practices expect that proper credit is given via citation. If you are unable to give datasets a CC0 waiver you may enter custom Terms of Use for datasets.

Yes, apply CC0 - "Public Domain Dedication"

No, do not apply CC0 - "Public Domain Dedication"

Terms of Use ?

If you are unable to use CC0 for datasets you are able to set custom terms of use. Here is an example of a [Data Usage Agreement](#) for datasets that have de-identified human subject data.

2. Add another open licence under "Terms of Use"

Licence (what's this?)

CC BY 4.0

CC BY 4.0

CC0

MIT

GPL

GPL 2.0+

GPL 3.0+

Apache 2.0

Figshare

License: Add a license

Choose a license: No license

No license

Year:

2020

Copyright Holders

Copyright 2020

Content:

CC0 1.0 Universal

CC-By Attribution 4.0 International

Code - Permissive:

MIT License

Apache License 2.0

BSD 2-Clause "Simplified" License

BSD 3-Clause "New"/"Revised" License

Code - Copyleft:

GNU General Public License (GPL) 3.0

GNU General Public License (GPL) 2.0

OSF

License

required

Access right *

- Open Access
- Embargoed Access
- Restricted Access
- Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

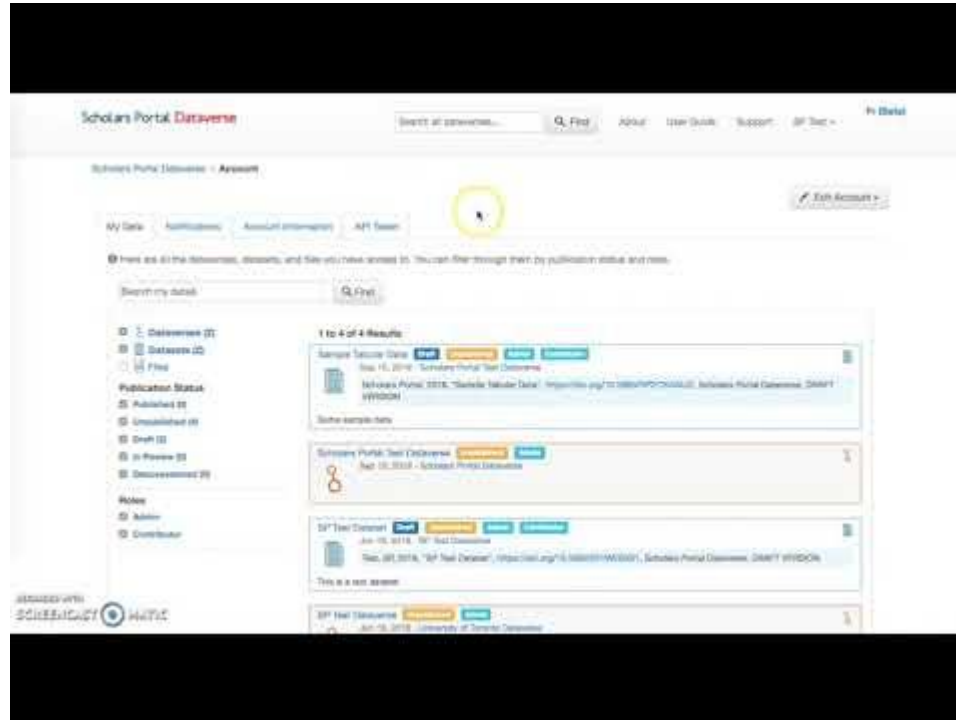
Zenodo

License *

Start typing a license name...

- Creative Commons Attribution 4.0 International**
- Creative Commons Attribution 1.0 Generic
- Creative Commons Attribution 2.0 Generic
- Creative Commons Attribution 3.0 Unported

Publishing datasets in Scholars Portal Dataverse



The screenshot displays the Scholars Portal Dataverse interface. At the top, the header includes the logo, a search bar, and navigation links for 'Home', 'User Guide', 'Support', and 'Sign Out'. Below the header, the user is logged in as 'Admin'. The main content area is titled 'My Data' and contains a search bar and a list of datasets. A yellow circle highlights the 'My Data' tab. The dataset list shows the following entries:

- Sample Salary Data** (Nov 10, 2014) - Secondary Portal Test Dataverse
- Scholars Portal 2014 "Market Value Data"** (September 10, 2014) - Scholars Portal Dataverse, Draft WOOD
- Scholars Portal Test Dataverse** (Sep 10, 2014) - Scholars Portal Dataverse
- SP Test Dataset** (Jun 10, 2014) - SP Test Dataverse
- SP Test Dataset** (Jun 10, 2014) - University of Toronto Libraries

At the bottom left, logos for 'Scholarly Commons' and 'HAC' are visible.

APRIL 15, 2021

Call for Applications: NDRIO-Portage COVID-19 Data Curation Funding

NDRIO and Portage are pleased to announce the availability of funding to support the curation of Canadian research data related to the COVID-19 pandemic. The purpose of this funding is to bring COVID-19 related data into timely compliance with the [FAIR Guiding Principles](#) (Findable, Accessible, Interoperable, and Reusable) for the management of research data, algorithms, tools and workflows.^[1] Ultimately, increasing the FAIRness of datasets makes them discoverable and re-usable by both humans and machines, enabling downstream research. As a secondary objective, this initiative aims to build capacity for research data management (RDM) in Canada by supporting the training of highly-qualified personnel (HQP) in data stewardship and curation.

About the COVID-19 Data Curation Funding Call

This initiative will contribute up to CAD \$336,000 toward the employment of [highly-qualified personnel](#) by [eligible applicants](#) for data stewardship and curation activities taking place between the date of the award and March 31, 2022. The maximum funding available per project is CAD \$42,000. This funding may only be used for the contracting of HQP for the stewardship or curation of data from research related to the COVID-19 pandemic. The amount of funding awarded to each recipient will be based on the scope and defined needs of each project as demonstrated through the application process.

Additionally, funding recipients will receive support from Portage in the form of an initial consultation with the Data Curation Team and with the possibility of group or individual training activities during the course of the project, as determined by the needs of recipients and in consultation with Portage.

<https://portagenetwork.ca/news/call-for-applications-ndrio-portage-covid-19-data-curation-funding/>

Key Takeaways

- Ensure you have the authority to share the data
- Choose a disciplinary or generalist repository
- Add descriptive metadata
- Share your data under an open licence such as Creative Commons
- Cite your data

Thank you!

Questions?

Mélanie Brunet, Copyright Services Librarian

melanie.brunet@uottawa.ca

Jeanette Hatherill, Scholarly Communication Librarian

jeanette.hatherill@uottawa.ca

Chantal Ripp, Research Data Management Librarian
(Interim)

chantal.ripp@uottawa.ca

Credits:

[Ulysses presentation template](#) by [SlidesCarnival](#), [CC BY 4.0](#)

Unless otherwise noted, icons used in this presentation are from the [Noun Project](#), used with permission through a [NounPro for Education royalty-free licence](#)

Additional Resources

- [FAQ: COVID-19 Rapid Response Data Sharing and Deposit Support](#), version 1.1, April 2020, Felicity Tayler and Chantal Ripp, uOttawa, [CC BY 4.0](#)
- Portage COVID-19 Working Group (2020, September 21). *Guide to COVID-19 Rapid Response Data Sharing and Deposit for Canadian Researchers* (Version 2). Zenodo. <http://doi.org/10.5281/zenodo.4270501>, [CC BY 4.0](#)
- RDA COVID-19 Working Group (2020). *Recommendations and Guidelines on data sharing*. Research Data Alliance, <https://doi.org/10.15497/rda00052>
- Teresa Scassa (2018). [Data Ownership](#). CIGI Papers No. 187, [CC BY-NC-ND 3.0](#)

Additional Resources

- Sensitive Data Expert Group. (2020, October 1). *Sensitive Data Toolkit for Researchers Part 3: Research Data Management Language for Informed Consent*. Zenodo. <http://doi.org/10.5281/zenodo.4107178>, [CC BY-NC 4.0](#)
- ARDC (2018). [FAIR Self Assessment Tool](#), Australian Research Data Commons
- Jones, Sarah, & Grootveld, Marjan. (2017, November). *How FAIR are your data?* Zenodo. <http://doi.org/10.5281/zenodo.1065991>, [CC BY 4.0](#)

Copyright ownership and data categories

Primary data:

- Data collected for your own purposes from experiment or research you have conducted
 - If copyright exists, you would likely be the owner (verify agreement or contract related to your research project to confirm)

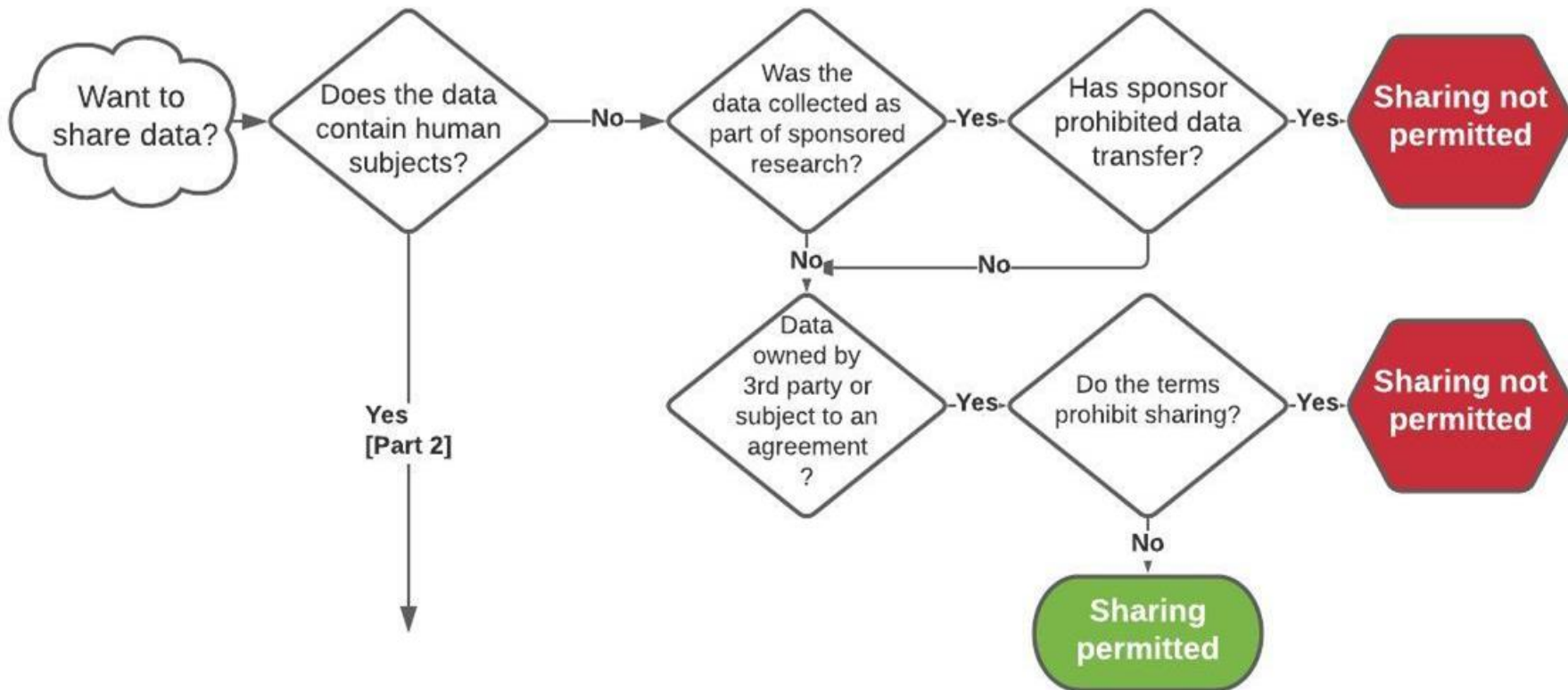
Secondary data:

- Data collected for other purposes from experiment or research conducted by others
 - If copyright exists, it is owned by others

Tertiary data:

- Synthesis of data from experiment or research conducted by others
 - Articles, reports, etc. written by others over which you do not own the copyright

Part 1 - Decision Tree: Can I share my data?



Adapted from: University of New Mexico, [Data Sharing Permitted Decision Tree](#).

Part 2 - Decision Tree: Can I share my data?

