

Recommended Repositories for COVID-19 Research Data

Prepared by the Portage Network, COVID-19 Working Group on behalf of the Canadian Association of Research Libraries (CARL)

Roger Reka (University of Windsor)
Minglu Wang (York University)
Erin Clary (Portage)
Beth Knazook (Portage)
Qian Zhang (University of Waterloo)
Kristi Thompson (Western University)
Lucia Costanzo (University of Guelph)
Kathy Szigeti (University of Waterloo)
Rebecca Dickson (COPPUL)
Mark Leggott (RDC-DRC)
Felicity Tayler (University of Ottawa)
Chantal Ripp (University of Ottawa)
Jane Fry (Carleton University)
Nick Rochlin (University of British Columbia)
Melanie Parlette-Stewart (Portage)

SEPTEMBER 2020

Portage Network
Canadian Association of Research Libraries
portage@carl-abrc.ca

www.carl-abrc.ca

portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

Table of Contents

Repository Options: Disciplinary or General Purpose?	2
Option 1: Use a Disciplinary Data Repository	2
<i>Recommended Disciplinary Repositories</i>	3
Option 2: Use a General-purpose Canadian or International Research Data Repository	5
<i>Recommended General Purpose Repositories</i>	5
Option 3: Choose your own Repository.....	6
Overview of Important Repository Features	7
<i>Things to Keep in Mind</i>	7
References	9

Recommended Repositories for COVID-19 Research Data

Sharing research data drives new research and informs policy and decision-making. During a pandemic, rapid access to the latest data informs public health response and helps save lives. Preserving these data for future use can also help to clarify the long-term economic, political, and cultural impacts of public health and policy decisions, and may inform the response to future pandemics. Use this document to help select a repository that will provide immediate and long-term access to your COVID-19 data.

Repository Options: Disciplinary or General Purpose?

Option 1: Use a Disciplinary Data Repository

Research data should be deposited in a repository that is recognized by other researchers in your community. Where such a repository is unavailable, or not suitable, you can deposit your research data in a generalist repository (see Option 2).

Using a disciplinary research data repository has its advantages:

- your research data are more visible to colleagues undertaking similar research
- the repository can be designed in a way to best support your type of data
- many research data repositories are curating collections of COVID-19 research data to enhance the discoverability of this type of data

When selecting a disciplinary research data repository, look for:

- suggestions or requirements outlined in funder guidelines (e.g., [CIHR](#))¹
- lists of recommended disciplinary data repositories maintained by various journals, including [Scientific Data](#) and [PLOS ONE](#).²
- community recognition (Which repositories are widely used in your field? These are the ones most likely to be accessed by your peers.)

¹ “CIHR Open Access Policy: Annex,” Canadian Institutes of Health Research, Last modified March 15, 2015, <https://cihr-irsc.gc.ca/e/46068.html#8>.

² Scientific Data, “Recommended Data Repositories,” Nature Research, accessed August 4, 2020, <https://www.nature.com/sdata/policies/repositories>; PLOS ONE, “Recommended Repositories.” PLOS, accessed August 4, 2020, <https://journals.plos.org/plosone/s/recommended-repositories>.

Recommended Disciplinary Repositories

This is a selected list of repositories well-suited to sharing COVID-19 data, but it is by no means complete. There are also excellent recommendations available in the [RDA COVID-19 Recommendations and Guidelines](#).³

- Public health & epidemiology
 - [OpenICPSR COVID-19 Data Repository](#)
 - Datasets [deposited](#) to OpenICPSR are not reviewed for disclosure risk, but a series of questions will help you assess whether the data can be public. Data that contain identifiable or sensitive information will be published as restricted access (servers are in the US).
 - A professional curation package that includes disclosure analysis is available for purchase. Contact ICPSR Acquisitions at deposit@icpsr.umich.edu for information.
- Social sciences
 - [OpenICPSR COVID-19 Data Repository](#)
 - [Qualitative Data Repository](#)
 - Consultation and curation is available for all deposits. [Access control](#) for sensitive data is possible (servers are in the US).⁴ There is a [service fee](#) for researchers affiliated with non-member institutions.⁵
- Software
 - [Software Heritage Archive](#)
 - [Zenodo \(with GitHub integration\)](#)
- Clinical data
 - [Vivli](#) for clinical trials data
 - Provides restricted access. [D-wise will waive fees to anonymize COVID-19 studies](#) made available through the Vivli platform, and Vivli will waive fees to share, archive and access all COVID-19 clinical trials data.⁶
 - [OpenNeuro](#) for MRI, MEG, EEG, iEEG, and ECoG
 - Open access repository. Data must be de-identified before sharing.
 - [EBRAINS](#) for raw or processed human brain data, models, analyses, methods, code, software
 - Open access repository. Data must be de-identified before sharing.

³ RDA COVID-19 Working Group, "RDA COVID-19 Recommendations and Guidelines on Data Sharing," June 30, 2020, <https://doi.org/10.15497/rda00052>.

⁴ "Human Participants – General Guidance," Qualitative Data Repository, accessed August 5, 2020, <https://qdr.syr.edu/guidance/human-participants>.

⁵ "Deposit Fees," Deposit, Qualitative Data Repository, accessed on August 5, 2020, <https://qdr.syr.edu/deposit#deposit-fees>.

⁶ "d-wise offers anonymization services to studies available on Vivli COVID-19 portal," Vivli, April 13, 2020, <https://vivli.org/vivli-covid-19-portal/>.

- [RSNA International COVID-19 Open Radiology Database](#) for chest imaging (CT and X-ray)
 - Open access repository. Images must be de-identified. RSNA has published a [de-identification protocol](#) and the Anonymizer software to guide this process.
- The [Immunology Database and Analysis Portal](#) (ImmPort) for cytometry and immunology
- Nucleic Acids & Omics
 - [International Nucleotide Sequence Database Collaboration \(INSDC\) archives](#) for genomics data (raw virus sequence data, gene expression data)
 - [NCBI GenBank](#) for genomics data (assembled and annotated genomes)
 - [Worldwide Protein Data Bank \(wwPDB\)](#) for structural data
 - See [RDA Guidelines](#) section 4.4 (page 28-36) for specific recommendations for virus and host genomics, structural data, proteomics, metabolomics, lipidomics.⁷

⁷ RDA COVID-19 Working Group, “RDA COVID-19 Recommendations and Guidelines on Data Sharing.”

Option 2: Use a General-purpose Canadian or International Research Data Repository

There are many generalist research data repositories available, including some provided by Canadian university libraries. These repositories provide a place to deposit data regardless of the research topic and may be better able to handle a wide variety of data types.

You may choose to deposit in a generalist data repository because:

- there is not a suitable disciplinary repository
- you wish to deposit analysis scripts and other supporting information with your data
- you'd like to share your data with a broader audience

Canadian repository options will allow you to keep your data on Canadian servers, take advantage of curation services at your institutions (where applicable), and make your data visible to broader data communities.

There is increasing interoperability and visibility between Canadian and international repositories. The Canadian Federated Research Data Repository (FRDR) harvests metadata from many other Canadian repositories, including all Scholars Portal Dataverse repositories, and shares it with [OpenAIRE](#) to make Canadian data discoverable internationally.⁸

Recommended General Purpose Repositories

- [Scholars Portal Dataverse](#) (Canadian)*
 - Dataverse allows restricted access, but servers are not encrypted. The repository should be used to share de-identified and non-confidential data only.
- [Federated Research Data Repository \(FRDR\)](#) (Canadian)*
 - FRDR is an open access repository. Data must be de-identified before sharing.
- [Zenodo](#) (International)
 - Recommended for the dedicated "[Coronavirus Disease Research Community](#)"⁹
 - Zenodo allows restricted access, but servers are not encrypted. The repository should be used to share de-identified and non-confidential data only.

*In 2019, Portage published a [Guide to Repository Options in Canada](#) that describes and compares Dataverse and FRDR in some detail. If you're having trouble choosing between these two options, we recommend looking at the section: "How do I decide which repository option to use?"¹⁰

⁸ "Explore," OpenAIRE, accessed August 5, 2020, <https://explore.openaire.eu/search/find>.

⁹ "Coronavirus Disease Research Community," Zenodo, accessed on August 5, 2020, <https://zenodo.org/communities/covid-19/?page=1&size=20>.

¹⁰ Portage Network, Compute Canada, Canadian Association of Research Libraries, and Scholars Portal, "Repository Options in Canada: A Portage Guide," June 1, 2019, <https://doi.org/10.5281/zenodo.3966348>.

Option 3: Choose your own Repository

We highlighted a few repositories in our recommendations above, but there are many more options available for different disciplines and purposes. You may have luck finding a relevant repository using the discovery tools below.

Please be aware that most of these tools list a wide variety of repositories, some of which do not offer long-term preservation or may not accept data deposits (they are publishing or discovery platforms only). It may be useful to keep in mind this [Overview of Important Repository Features](#) while browsing these resources:

- This [Generalist Repository Comparison Chart](#) gives an overview of major international generalist repositories.¹¹
- The [Repository Finder](#) tool queries the [re3data registry of research repositories](#) to help researchers find appropriate repository options.¹²
- You can directly [browse re3data by subject](#), then use filters to further narrow your results.¹³ The ‘data upload’ and ‘data upload restrictions’ filters can help you eliminate repositories that don’t accept new deposits or those that limit who can deposit (e.g., limited to institutional membership).
- [FAIRsharing](#) may be used to find various standards and databases in your research area.¹⁴ Be aware that the databases category includes both repositories that can ingest new data deposits, and other resources that catalog and host data, but which may not accept direct data deposits from researchers.
 - The FAIRsharing team has created a [collection of standards and databases](#) that are appropriate for use in the COVID-19 pandemic.¹⁵
 - The Research Data Alliance COVID-19 Working Group has also curated a [collection of databases, policies, and standards](#) for researchers and policymakers.¹⁶

¹¹ Shelley Stall, Maryann E. Martone, Ishwar Chandramouliswaran, Mercè Crosas, Lisa Federer, Julian Gautier, Mark Hahnel, *et al.* “Generalist Repository Comparison Chart,” July 15, 2020, <http://doi.org/10.5281/zenodo.3946720>.

¹² DataCite, and American Geophysical Union, “Repository Finder,” DataCite, accessed August 4, 2020, <https://repositoryfinder.datacite.org/>; Berlin School of Library and Information Science, DataCite, Helmholtz Open Science Office, KIT Library, and Purdue University Libraries, “Registry of Research Data Repositories,” re3data, accessed August 4, 2020, <https://doi.org/10.17616/R3D>.

¹³ “Browse by Subject,” re3data, accessed on August 4, 2020, <https://www.re3data.org/browse/by-subject/>.

¹⁴ “Standards, Databases, Policies,” FAIRsharing, accessed August 4, 2020, <https://www.FAIRsharing.org>.

¹⁵ “COVID-19 Resources,” FAIRsharing, accessed August 5, 2020, <https://fairsharing.org/collection/COVID19Resources>.

¹⁶ RDA COVID-19 Working Group, “RDA Covid-19 WG Resources,” FAIRsharing, accessed August 5, 2020, <https://fairsharing.org/collection/RDACovid19WG>.

Overview of Important Repository Features

A good repository will do more than simply make your data accessible; it will ensure its trustworthiness. To facilitate this, repositories must have reliable infrastructure and good governance, policies that support their community (both depositors and users), and they should be able to demonstrate responsible data stewardship. The following recommendations are adapted from the widely-endorsed [TRUST Principles for Digital Repositories](#).¹⁷

- The repository has the ability to assign **unique identifiers** to datasets (e.g., [DOIs](#) or accession numbers)¹⁸
 - *Why is this important?* This will allow you and others to cite your datasets, linking your data to your research outputs and ensuring your data is discoverable.
- The repository has clear **terms of use** and **commits to the preservation of digital holdings**
 - *Why is this important?* Not all repositories support preservation. If you want to ensure your data is retained for the long term, you should choose a repository that has a clear policy on preservation.
- The repository provides **stewardship of the data holdings**, including technical validation, documentation, and quality control
 - *Why is this important?* Both depositors and users need to have confidence in the enduring quality of the data in a repository.
- The repository has a mechanism for **user support**
 - *Why is this important?* Repositories should be transparent and easily understood by the end user, monitoring changing expectations and responding to them.
- The repository is **recognized within its research community**, and supports the designated community's metadata and curation standards
 - *Why is this important?* Where do you go to look for data in your field? Researchers should be able to find your data in the places they would expect to find it.

Things to Keep in Mind

- What types of research outputs do you need to archive?
 - [RDA Guidelines](#) recommend that researchers archive situational data, analytical models, scientific findings, and reports used in decision-making.¹⁹ Not all repositories accept all these types of materials. You can highlight related resources in your README or other documentation, and many repositories have a 'related resources' field that will allow you to link directly to external resources.
- Where will your data be stored? Is it important for the data to be hosted within Canada?
 - If you have privacy or security concerns, you will want to consider a repository's storage location alongside its other features.
- Can all your data be shared? Do you need a repository that can restrict access?

¹⁷ Dawai Lin, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giarretta, Marisa De Giusti, *et al*, "The TRUST Principles for Digital Repositories," *Scientific Data* 7, 144 (2020), <https://doi.org/10.1038/s41597-020-0486-7>.

¹⁸ "The DOI System, ISO 26324," DOI. Last modified February 4, 2020, <https://www.doi.org>.

¹⁹ RDA COVID-19 Working Group, "RDA COVID-19 Recommendations and Guidelines on Data Sharing."

- If you have questions about how open you can make your data, check out the [Can I Share My Data?](#) decision tree and [De-identification Guidance](#) before proceeding.²⁰

²⁰ Portage COVID-19 Working Group, “Can I Share My Data?” September 25, 2020, <https://doi.org/10.5281/zenodo.4041661>; Portage COVID-19 Working Group, “De-identification Guidance.” September 25, 2020, <https://doi.org/10.5281/zenodo.4042022>.

Repositories

1. EBRAINS <https://ebrains.eu/services/data-knowledge/share-data>
2. Federated Research Data Repository (FRDR) <https://www.frdr-dfdr.ca/repo/>
3. Immunology Database and Analysis Portal (ImmPort) <https://www.immport.org/resources/dataTemplates>
4. International Nucleotide Sequence Database Collaboration (INSDC) archives <http://www.insdc.org/>
5. NCBI GenBank <https://www.ncbi.nlm.nih.gov/genbank/>
6. OpenICPSR COVID-19 Data Repository <https://www.openicpsr.org/openicpsr/covid19>
7. OpenNeuro <https://openneuro.org/>
8. Qualitative Data Repository <https://qdr.syr.edu/deposit/process>
9. RSNA International COVID-19 Open Radiology Database <https://www.rsna.org/covid-19/COVID-19-RICORD>
10. Scholars Portal Dataverse <https://dataverse.scholarsportal.info/>
11. Software Heritage Archive <https://www.softwareheritage.org/save-and-reference-research-software/>
12. Vivli <https://vivli.org/>
13. Worldwide Protein Data Bank (wwPDB) <http://www.wwpdb.org/>
14. Zenodo <https://zenodo.org/>

References

1. Berlin School of Library and Information Science, DataCite, Helmholtz Open Science Office, KIT Library, and Purdue University Libraries. "Registry of Research Data Repositories." re3data. Accessed August 4, 2020. <https://doi.org/10.17616/R3D>.
2. Canadian Institutes of Health Research. "CIHR Open Access Policy: Annex." Last modified March 15, 2015. <https://cihr-irsc.gc.ca/e/46068.html#8>.
3. DataCite, and American Geophysical Union. "Repository Finder." DataCite. Accessed August 4, 2020. <https://repositoryfinder.datacite.org/>.
4. DOI. "The DOI System, ISO 26324." Last modified February 4, 2020. <https://www.doi.org>.
5. FAIRsharing. "Standards, Databases, Policies." Accessed August 4, 2020. <https://www.FAIRsharing.org>.
6. FAIRsharing. "COVID-19 Resources." Accessed August 5, 2020. <https://fairsharing.org/collection/COVID19Resources>.
7. GitHub. "GitHub Guides: Making Your Code Citable." Last modified October 2016. Accessed August 5, 2020. <https://guides.github.com/activities/citable-code/>.

8. Lin, Dawai, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, *et al.* "The TRUST Principles for Digital Repositories," *Scientific Data* 7, 144 (2020). <https://doi.org/10.1038/s41597-020-0486-7>.
9. OpenAIRE. "Explore." Accessed August 5, 2020. <https://explore.openaire.eu/search/find>.
10. PLOS ONE. "Recommended Repositories." PLOS. Accessed August 4, 2020. <https://journals.plos.org/plosone/s/recommended-repositories>.
11. Portage COVID-19 Working Group. "Can I Share My Data?" September 25, 2020. <https://doi.org/10.5281/zenodo.4041660>.
12. Portage COVID-19 Working Group. "De-identification Guidance." September 25, 2020, <https://doi.org/10.5281/zenodo.4042022>.
13. Portage Network, Compute Canada, Canadian Association of Research Libraries, and Scholars Portal. "Repository Options in Canada: A Portage Guide." June 1, 2019. <https://doi.org/10.5281/zenodo.3966348>.
14. Portage Network. "How to Manage Your Data. Contacts at Your Organization." <https://portagenetwork.ca/planning-managing-data/contacts-at-your-organization>.
15. Qualitative Data Repository. "Deposit Fees." Accessed on August 5, 2020. <https://qdr.syr.edu/deposit#deposit-fees>.
16. Qualitative Data Repository. "Human Participants – General Guidance." Accessed August 5, 2020. <https://qdr.syr.edu/guidance/human-participants>.
17. RDA COVID-19 Working Group. "RDA Covid-19 WG Resources." FAIRsharing. Accessed August 5, 2020. <https://fairsharing.org/collection/RDACovid19WG>.
18. RDA COVID-19 Working Group. "RDA COVID-19 Recommendations and Guidelines on Data Sharing." June 30, 2020. <https://doi.org/10.15497/rda00052>.
19. re3data. "Browse by Subject." Accessed on August 4, 2020. <https://www.re3data.org/browse/by-subject/>.
20. Scientific Data. "Recommended Data Repositories." Nature Research. Accessed August 4, 2020. <https://www.nature.com/sdata/policies/repositories>.
21. Stall, Shelley, Maryann E. Martone, Ishwar Chandramouliswaran, Mercè Crosas, Lisa Federer, Julian Gautier, Mark Hahnel, *et al.* "Generalist Repository Comparison Chart". July 15, 2020. <http://doi.org/10.5281/zenodo.3946720>.
22. Vivli. "d-wise offers anonymization services to studies available on Vivli COVID-19 portal." April 13, 2020. <https://vivli.org/vivli-covid-19-portal/>.
23. Zenodo. "Coronavirus Disease Research Community." Accessed on August 5, 2020. <https://zenodo.org/communities/covid-19/?page=1&size=20>.