# Documentation and Supporting Materials Required for Deposit

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

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

SEPTEMBER 2020

# Table of Contents

# Documentation and Supporting Materials Required for Deposit

The speed with which COVID-19 research is being produced and shared only increases the need for good documentation and description. It may be necessary to make available the results of preliminary trials or unprocessed datasets, which may be difficult for fellow researchers to interpret without contextual information. To ensure your data are independently understandable, please consider providing:

- a codebook and/or a README file,
- a description of the methods you used to collect and/or process the data,
- any associated code, scripts, or syntax files you used to process or analyze your data, and
- a data use agreement if any data will be restricted.

## Create a README File

When creating a codebook and/or README file, please consider the following:

- Include a point of contact.
- List any restrictions on secondary use of your data.
- For quantitative datasets, define all variables and allowable values. When applicable, include units of measure, and define the code you used for missing or null values.
- Include a brief description of your study, the methods you used to collect your data, and any steps you took to anonymize or otherwise process your data.
- If you removed variables from your raw dataset to create a public use copy for archiving, include a list of the variables that were removed so the changes made to your raw dataset are transparent. You might also choose to provide summary statistics for any variable that was removed.
- The names of equipment or instruments used to collect data, and software or statistical packages that were used to process the data should be listed. If possible, include the version of software you used.
- If your file formats are not plain text, include a recommendation for software that can be used to view or analyze the files.

Further guidance is available in UBC's Quick Guide: Creating a README for your dataset and Cornell University's Guide to writing "readme" style metadata.[1] Cornell has also published a README template that you are free to download and modify for your own dataset.[2]

---

[1] Doug Brigham, "Creating a README for Your Dataset: Quick Guide, Version 1.0," The University of British Columbia Library. April 2020. https://researchdata-06oct2014.sites.olt.ubc.ca/files/2020/04/QuickGuide_UBC_readme_v1.0_20200427.pdf. See also, "Guide to writing 'readme' style metadata," Cornell University, accessed August 17, 2020, https://data.research.cornell.edu/content/readme.

[2] "Author-Dataset-ReadmeTemplate.txt," Cornell University, accessed on August 17, 2020, https://cornell.app.box.com/v/ReadmeTemplate.

## Include Recommended Supplementary Materials

Please consider depositing the following documents as well, or link to them from your codebook or README if they have been deposited elsewhere:

- A copy of your Data Management Plan.
- An unsigned copy of the consent form you provided to study participants. This may be archived alongside your data, or you may provide a copy to the data repository if your data will be screened for disclosure risk before they are archived.
- Consider archiving supplemental materials that provide context to your dataset, for example, study protocols, clinical study reports, and statistical analysis plans.
- Plan to include links to any published article(s) or related resources that report on the results of your analyses. You may be able to add links to your publication(s) later, if you have deposited your data first.

## Add Repository Metadata

Another way to provide context to your data is to add rich metadata when you deposit it. Some disciplinary repositories have required standards you must adhere to, but many repositories require only basic information in order to publish, such as title, contact author, and a choice of license. To increase the findability and usability of your data, plan to include a descriptive title and keywords. A robust description will also help other researchers understand your data and will be especially important if access to your dataset is restricted. It may be helpful to think of the dataset description as you would an article abstract. The description may include high-level information about the dataset such as the data collection methodology, steps you took to process or analyze the data, the results of your analyses, and other potential uses for the data.

## Specify Data Use Agreements

If you will be depositing restricted materials, any restrictions on who can access your data for secondary use should be made explicit. You may negotiate a data use agreement with the repository that accepts your data, but it is also advisable to clearly indicate any restrictions in your README file. Repository terms of use tend to be fairly simple (e.g. https://www.immport.org/agreement).[3] Dataverse has provided some more complex sample data use agreement language that you may want to consult if you will be providing de-identified human subject data: Harvard Dataverse Sample Data Usage Agreement.[4]

---

[3] "User Agreement for the NIAID Immunology Database and Analysis Portal (ImmPort)," ImmPort, accessed on August 19, 2020, https://www.immport.org/agreement.

[4] "Harvard Dataverse – Sample Data Use Agreement," Dataverse, accessed on August 17, 2020, http://best-practices.dataverse.org/harvard-policies/sample-dua.html.

# References

1. Brigham, Doug. "Creating a README for Your Dataset: Quick Guide, Version 1.0." The University of British Columbia Library. April 2020. https://researchdata-06oct2014.sites.olt.ubc.ca/files/2020/04/QuickGuide_UBC_readme_v1.0_20200427.pdf.
2. Cornell University. "Guide to writing 'readme' style metadata." Accessed August 17, 2020. https://data.research.cornell.edu/content/readme.
3. Cornell University. "Author-Dataset-ReadmeTemplate.txt." Accessed on August 17, 2020. https://cornell.app.box.com/v/ReadmeTemplate.
4. Dataverse. "Harvard Dataverse – Sample Data Use Agreement." Accessed on August 17, 2020. http://best-practices.dataverse.org/harvard-policies/sample-dua.html.
5. ImmPort. "User Agreement for the NIAID Immunology Database and Analysis Portal (ImmPort)." Accessed on August 19, 2020. https://www.immport.org/agreement.