Portage Network: CRDCN Template for Research Data Centres and External Analysis

Data Collection

Which RDC datasets will be used in the research?

Example Answer:

The data source(s) for this project is/are the <<INSERT NAME OF SURVEYS/ADMINISTRATIVE RECORDS APPROVED>>. The current version(s) is/are: <<Record number>>.

Guidance:

The record number is available on Statistics Canada's website which can be accessed directly, or through our website: crdcn.org/data. E.g. Aboriginal People's Survey 2017 Record number:3250 https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3250

Please describe the collection process for the supplemental or external data that will be part of your project.

Guidance:

External or Supplemental data are the data used for your research project that are not provided to you by Statistics Canada through the Research Data Centre program.

What file formats will the supplementary data be collected and processed in? Will these formats permit sharing and long-term access to the data? How will you structure, name and version these files in a way easily understood by others?

Documentation and Metadata

What will you do to ensure that your research data contributions (syntax, output etc…) in your RDC project folder and (if applicable) your external analysis are properly documented, organized and accessible?

Guidance:

Resources are available on the CRDCN website to help. A recommendation from CRDCN on how to document your research contributions can be found here. For ideas on how to properly curate reproducible research, you can go here: https://labordynamicsinstitute.github.io/replication-tutorial-2019/#/

How will you make sure that the syntax archived in your project folder (and if applicable that created for your external analysis) is created consistently throughout your project?

Guidance:

Syntax: Any code used by the researcher to transform the raw data into the research results. This most commonly includes, but is not limited to, .do (Stata) files, .sas (SAS) files, and .r (R) R code.

Please provide the information about the availability of the metadata for your project here (both the RDC data and your external data). Some metadata for RDC datasets is available by contacting the RDC analyst.

How will you ensure that the external/supplemental data are easily understood and correctly documented (including metadata)?

Guidance:

For a good starter resource on metadata see: https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/.

Storage and Backup

What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

Guidance:

Because of the structure of the agreements under which supplemental data are brought into the RDC we highly recommend a parallel storage and backup to simplify sharing of these research data. Note that "data" here refers not only to the raw data, but to any and all data generated in the course of conducting the research.

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

How will the research team and other collaborators access, modify, and contribute data throughout the project?

Preservation

Will you deposit your syntax and other research data in a repository to preserve your files? Please describe your intended preservation of all research data here, noting how you will deal with any privacy concerns related to your supplemental/external data:

Sharing and Reuse

Outside of the data sharing/reuse that happens automatically within your project folder, what data will you be sharing, where, and in what form (e.g. raw, processed, analyzed, final)?

Guidance:

Consider also what file-format you will use. Will this file format be useable in the future? Is it proprietary?

What type of end-user license will these shared data fall under?

What steps will you take to help the research community know that these data exist?

Responsibilities and Resources

For the supplemental/external data, identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.

For the supplemental/external data, how will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

For the supplemental/external data, what resources will you require to implement your data management plan for all your research Data (i.e. RDC data and external/supplemental)? What do you estimate the overall cost for data management to be?

Guidance:

A tool provided by OpenAIRE can help researchers estimate the cost of research data management: https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs.

Ethics and Legal Compliance

If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

Guidance:

Consider where, how, and to whom sensitive data with acknowledged long-term value should be made available, and how long it should be archived. Decisions should align with Research Ethics Board requirements. Methods used to share data will be dependent on the type, size, complexity and degree of sensitivity of data. Outline problems anticipated in sharing data, along with causes and possible measures to mitigate these. Problems may include confidentiality, lack of consent agreements, or concerns about Intellectual Property Rights, among others.

Reused from: Digital Curation Centre. (2013). Checklist for a Data Management Plan. v.4.0. Restrictions can be imposed by limiting physical access to storage devices, placing data on computers with no access to the Internet, through password protection, and by encrypting files. Sensitive data should never be shared via email or cloud storage services such as Dropbox

If applicable, what strategies will you undertake to address secondary uses of sensitive data?

Guidance:

Obtaining the appropriate consent from research participants is an important step in assuring Research Ethics Boards that the data may be shared with researchers outside your project. The consent statement may identify certain conditions clarifying the uses of the data by other researchers. For example, it may stipulate that the data will only be shared for non-profit research purposes or that the data will not be linked with personally identified data from other sources. Read more about data security: UK Data Archive.

How will you manage legal, ethical, and intellectual property issues?

Guidance:

Compliance with privacy legislation and laws that may impose content restrictions in the data should be discussed with your institution's privacy officer or research services office. Research Ethics Boards are central to the research process.

Include here a description concerning ownership, licensing, and intellectual property rights of the data. Terms of reuse must be clearly stated, in line with the relevant legal and ethical requirements where applicable (e.g., subject consent, permissions, restrictions, etc.).

If you feel there are any other legal or ethical requirements for your project please describe them here: