Data Management Plan Template:
Accessing Data from Research Data Centres (CRDCN)

Abstract

The Canadian Research Data Centre Network (CRDCN) template summarizes the data management that is conducted by Statistics Canada and the CRDCN on behalf of researchers. While there are some advantages to working inside the RDC for data management, there is also a substantial drawback: RDC data can never be deposited in a repository in accordance with the recommended best practices for research data management. Because of this, researchers should be mindful of other options to engage in best practices. In addition to ensuring that the RDC project folder is well documented, and consistent with the research output, researchers should curate a supporting data deposit at a recognized repository in their discipline or within the Federated Research Data Repository (FRDR) containing metadata, syntax (code that produces a statistical output), and any other supporting material for the research project.

This template is for researchers who are doing RDC work using Statistics Canada data available in the RDC *only* (i.e. there is no supplemental data, public use statistics, or any other information that complements the RDC work). If your work is being conducted in the RDC in concert with other data that you either intend to bring into the RDC or work on outside the RDC in parallel to your RDC work, then the RDC and External Analysis template should be completed.

Administrative Details

Version:

Version	Date	Changes
1.0	2021-04-09	Formatted for inaugural publication.

Data Collection

All research conducted in the Research Data Centres (hereafter RDC), using Statistics Canada data exclusively, is secondary in nature. There is no data collection involved in this project. These data are owned and maintained by Statistics Canada with storage and access provided by the Canadian Research Data Centres Network.

Raw data in the RDC are stored in multiple formats including but not limited to .SAS (SAS), .dta (STATA), and .shp (shapefiles) as appropriate. The availability of StatTransfer™ software within the RDCs and continued management by Statistics Canada will ensure that the data will be accessible indefinitely should the file formats currently in use become obsolete.

The data provided by Statistics Canada are assigned unique identifiers which can be used to identify the data in any research output.

Which RDC datasets will be used in the research?

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).

Example: 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>>.

Documentation and Metadata

Documentation provided by Statistics Canada in the RDC is available to any data-users. This documentation is freely available to those with approved projects, and contains information about the sample selection process, a copy of the questionnaire, and a codebook.

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?

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.

How will you make sure that the syntax archived in your project folder is created consistently throughout your project?

Some metadata is available by contacting the RDC analyst. Is the metadata for the data to be used in your analysis available outside of the RDC? Please provide the information about the availability of the metadata for your project here.

You can consult your analyst to learn more about the availability of metadata for your proposed dataset. In some cases, the codebooks contain confidential information (quantiles with small numbers of individuals identified etc.). and cannot be made available.

Storage and Backup

Data storage is managed by the CRDCN in partnership with Statistics Canada on Servers located across the network. The current policy of the CRDCN is to store project data (syntax, releases, and anything else stored in the project folder) for ten years. Data are backed up on site and accessible through a highly secured network from any of the other RDC locations.

Preservation

The work conducted in the RDC for this project is kept based on the Contract ID provided by the RDC program which can be used by anyone on the project team to retrieve the code and supporting documents for a period of 10 years as described above in "Storage and Backup". Raw data that is the property of Statistics Canada, i.e. RDC data are permanently stored by Statistics Canada, but can never be released to the researcher. Researchers can also preserve all user-generated RDC research data that meets the criteria for release through a vetting request via a repository such as FRDR (though it is again emphasized that the raw RDC data cannot be shared). Best practices for reproducible work require indefinite preservation of research data (so in the case of RDC research, this means metadata, syntax, methodology).

Will you deposit your syntax and other research data in a repository to host your syntax files publicly? If so, please describe here.

Is there any other preservation that will be done as part of this research project? If so, please describe here.

Sharing and Reuse

Because the Statistics Canada Microdata files are collected under assurances of confidentiality and are owned and controlled by Statistics Canada, they cannot be shared by any member of the research team.

Access to the data in the RDCs is governed by the CRDCN's Access and Fee-for-service policy in <u>English</u> or <u>French</u>. The policy provides free access to university-based researchers who are network members and provides access to others on a cost-recovery basis.

The CRDCN and Statistics Canada promote their data holdings through social media and their respective websites. In addition, CRDCN data are required to be cited in any and all publications with the Statistics Canada Record Number so that readers are able to find the data. In addition, all publications using RDC data should include the RDC contract ID so that potential users can find information on the original contract. This information is available on the CRDCN website (crdcn.org/publications).

If you feel there are any additional sharing and reuse concerns related to your project please describe them here.

Responsibilities and Resources

The CRDCN and Statistics Canada will maintain the research data even if the researcher leaves their organization.

CRDCN enjoys the support of CIHR, SSHRC and CFI as well as receiving funds from the partner universities. There is no charge to the users of the RDCs for the data management conducted under the auspices of CRDCN and Statistics Canada as described within this DMP.

CRDCN does not employ consistency checking to ensure that the code provided alongside requests for research results to be released from the secure facility truly creates the output as requested. The responsibility for ensuring that the code and documents describing their use work as intended and are clear to other users who might access them lies with the researchers in the RDC. The CRDCN has a mechanism to ensure that the code is saved alongside all of the research output used to support the conclusions of any published works.

In addition to the data management employed by Statistics Canada, it is possible for researchers to have research output that does not contain confidential data, including tables, syntax and other information, released from the RDC where it could be curated in a repository of the researcher's choosing as described in question 5. If you plan to do any supplemental storage or curation of your research data, please comment on where the responsibility for curation and maintenance of the archive resides.

Will any resources be required for this curation and maintenance? If so, please estimate the overall data management costs.

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

Any users of the RDC must be 'deemed employees' of Statistics Canada. To become a deemed employee, the Treasury Board mandates a security clearance process including a criminal background check, credit check and fingerprinting. Approval for access to data requires a peer-review process of a research proposal and an institutional review at Statistics Canada. In cases where a researcher's scholarly work has been assessed through the tenure review process, they are considered peer-review pre-approved and only the institutional review is required.

Once a researcher is granted access to the RDC they must take an Oath of Secrecy – promising never to disclose confidential data. Criminal penalties can apply under the Statistics Act for violations of this oath.

Intellectual property for work done within the RDC becomes property of Statistics Canada including code used to manipulate data. The collection and dissemination of, and access to, confidential microdata is conducted under the Statistics Act and complies with all legal requirements. The confidential microdata for this project cannot be shared, posted, or copied. Access to the data is available through the RDC program. More information on how to access data is available here in <u>English</u> or <u>French</u>.

In general, research ethics clearance is not required for research conducted in the RDC. A statement from the CRDCN on the topic is available here in <u>English</u> or <u>French</u>.

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