

# Data Management Plan Template: Research Data Centres and External Analysis (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 *and* research data that they have either brought into the RDC “supplemental data” or are analyzing in parallel to their work in the RDC (such as mixed-methods) or public use statistics that compliment the RDC work (hereafter: *external data*). Researchers should be aware that any data brought into the RDC will be stored alongside the rest of their project material subject to the information management protocols from Statistics Canada. This is a free, relatively straightforward, process and researchers can obtain more information by talking to their RDC analyst.

If your work is being conducted in the RDC using only data provided through the RDC program then the RDC-only template should be completed and not this template.

## Administrative Details

**Template Author(s):** Grant Gibson, Canadian Research Data Centre Network (CRDCN)

**Published:** April 9, 2021

**DOI:** [10.5281/zenodo.4667839](https://doi.org/10.5281/zenodo.4667839)

**Contact:** Portage Network - [portage@engagedri.ca](mailto:portage@engagedri.ca), [portagenetwork.ca](http://portagenetwork.ca)

**License:** [Attribution-NonCommercial 4.0 International \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)



*Data Management Plan Template:  
Research Data Centres and External Analysis (CRDCN)*

**Version:**

<b>Version</b>	<b>Date</b>	<b>Changes</b>
1.0	2021-04-09	Formatted for inaugural publication.

## **Data Collection**

*All research conducted in the Research Data Centres (hereafter RDC) is secondary in nature. There is no data collection involved in this portion of the 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. Researchers can bring data into the RDCs (these will be called “supplemental data”). When they do, they are stored alongside all of the other research products related to that contract from the RDC and archived.*

### **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](http://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>>.

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

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**

*Documentation provided by Statistics Canada in the RDC will be available to any potential future users of these data. 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. Researchers should also think about how the metadata for their external data can be provided to other researchers. Best practices require that there be coordination between the internal and external data management. How to best manage this will depend on the nature of the external data.*

**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?**

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?**

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)?**

For a good starter resource on metadata see:

<https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/>.

## **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. These data are backed up on site and accessible through a highly secured network from any of the other RDC locations. Raw data related to the research project are stored in perpetuity by Statistics Canada.*

*For external research data, storage and backup are solely the responsibility of the researcher. Please consider the following questions as they relate to external data. These questions should also be considered for supplemental data if you plan to do parallel storage and backup of these data.*

**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?**

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

*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 is 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). In addition to this preservation for the RDC work, the external data (and related syntax, metadata and methodology) should be preserved also.*

**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

*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 [English](#) or [French](#). 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 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](http://crdcn.org/publications)).*

*For your supplemental/external data, please answer the following questions aimed to satisfy the [FAIR principles](#).*

**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)?**

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**

*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.*

*Researchers should consider how to manage their external research data and should think about who on the project team will have responsibility for managing the research data and what resources might be required to do so. Where possible, the research data from within the RDC should be managed in a way that is coordinated with the external research data management.*

*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 the Preservation section. If you plan to do any supplemental storage or curation of your research data (either the user-generated research data from the RDC or the external/supplemental data), please comment on where the responsibility for curation and maintenance of this archive resides.*

**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?**

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 exclusively through Statistics Canada and the RDC program. More information on how to access data is available here in [English](#) or [French](#).*

*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 [English](#) or [French](#).*

*Please respond to the following ethical compliance questions as they relate to your external/supplemental data. If your project underwent research ethics review at your institution, you can summarize the submission instead of answering these questions.*



**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?**

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?**

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?**

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.**