

## Data Proposal Requirements

### General Proposal Guidelines

- Any proposal for submission must have sufficient details to determine the feasibility of the project.
- The purpose of access to confidential microdata is to support model-based inferential statistical modelling. Research proposals must be analytic with a focus on modelling.
- Researchers are expected to use [publicly available data tables](#) or [real time remote access \(RTRA\)](#) where possible to support analytic modeling outputs.
- This environment is not intended for descriptive-only analyses or large tabular releases. Descriptive statistics (e.g., cross-tabulations, summary/descriptive statistics) should be limited to supporting inferential modeling.
- If you wish to obtain custom tabulations, please contact [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.ca).
- All researchers who intend to include Indigenous Peoples (First Nations, Métis or Inuit), in their study population are expected to engage with Indigenous organizations, communities or researchers. Resources have been developed by First Nations, Métis and Inuit organizations that explain ethical considerations, the role of research in Indigenous self-determination, and the imperative for respectful research practices. *Please refer to the following document for guidance, [Working-with-FN-Inuit-Metis-data\\_EN.pdf](#)*
- Refer to these additional guidelines for proposals requesting the following data:
  - [Census and National Household Survey \(NHS\)](#)
  - [Longitudinal and International Study of Adults \(LISA\)](#)
  - [Survey of Household Spending \(SHS\)](#)

### Proposal Template

**It is strongly recommended that you review your proposal with your local Statistics Canada Data Access Analyst for feedback prior to submission.** RDC and SDAC contact information is listed at: [Contact information for the Data Access Division](#), under the section “Research Data Centres”.

The project proposal should be a maximum of ten pages and must follow the headings in the template below. Missing information or insufficient details may result in delays in processing your application.

## RESEARCH PROPOSAL FOR DATA ACCESS

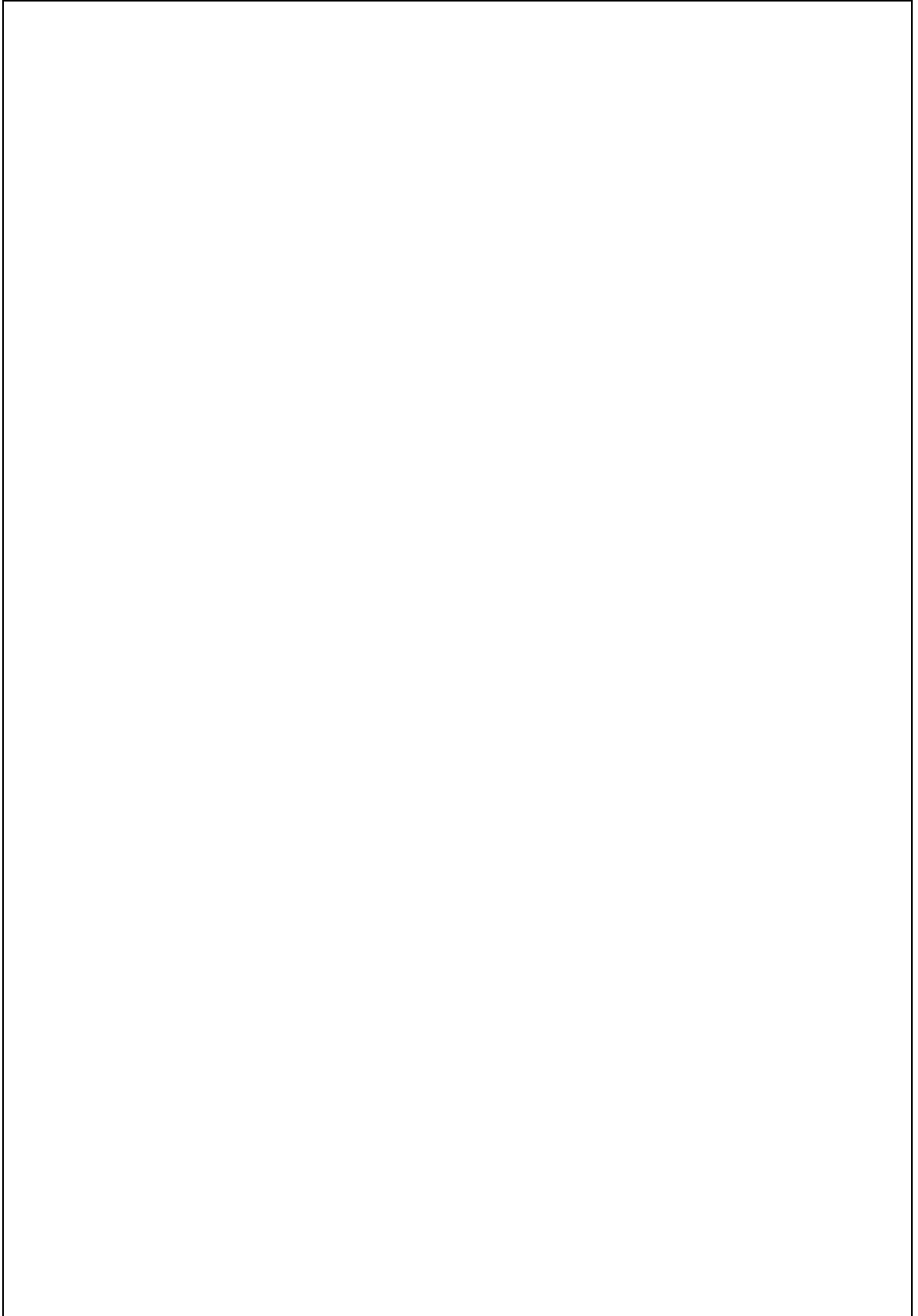
Principal Investigator & Affiliation	
Co-investigator(s) & Affiliation(s)	
<a href="#">Access Location</a> *	

**1. Project title:**

--

**2. Project description, rationale, and objectives:**

- a) Provide a summary of the current literature and identify the knowledge gap(s) to be addressed (approximately 1 page).



b) Clearly identify the specific questions or objectives of the project:

- Enumerate the specific effects and/or objectives that are being evaluated.
- Please provide a description of the engagement undertaken for this project, including any Indigenous partners, researchers or collaborators.
- If this project impacts First Nations people, Métis or Inuit explain the benefits for Indigenous Peoples and communities and identify any possible harm along with risk mitigation strategies.



**3. Data requirements:**

- a) Justification for using microdata (as opposed to alternative data access methods such as Public Use Microdata Files (PUMFs), Real Time Remote Access (RTRA), or custom tabulations).

- b) Dataset(s) required: Please specify how each dataset will be used.

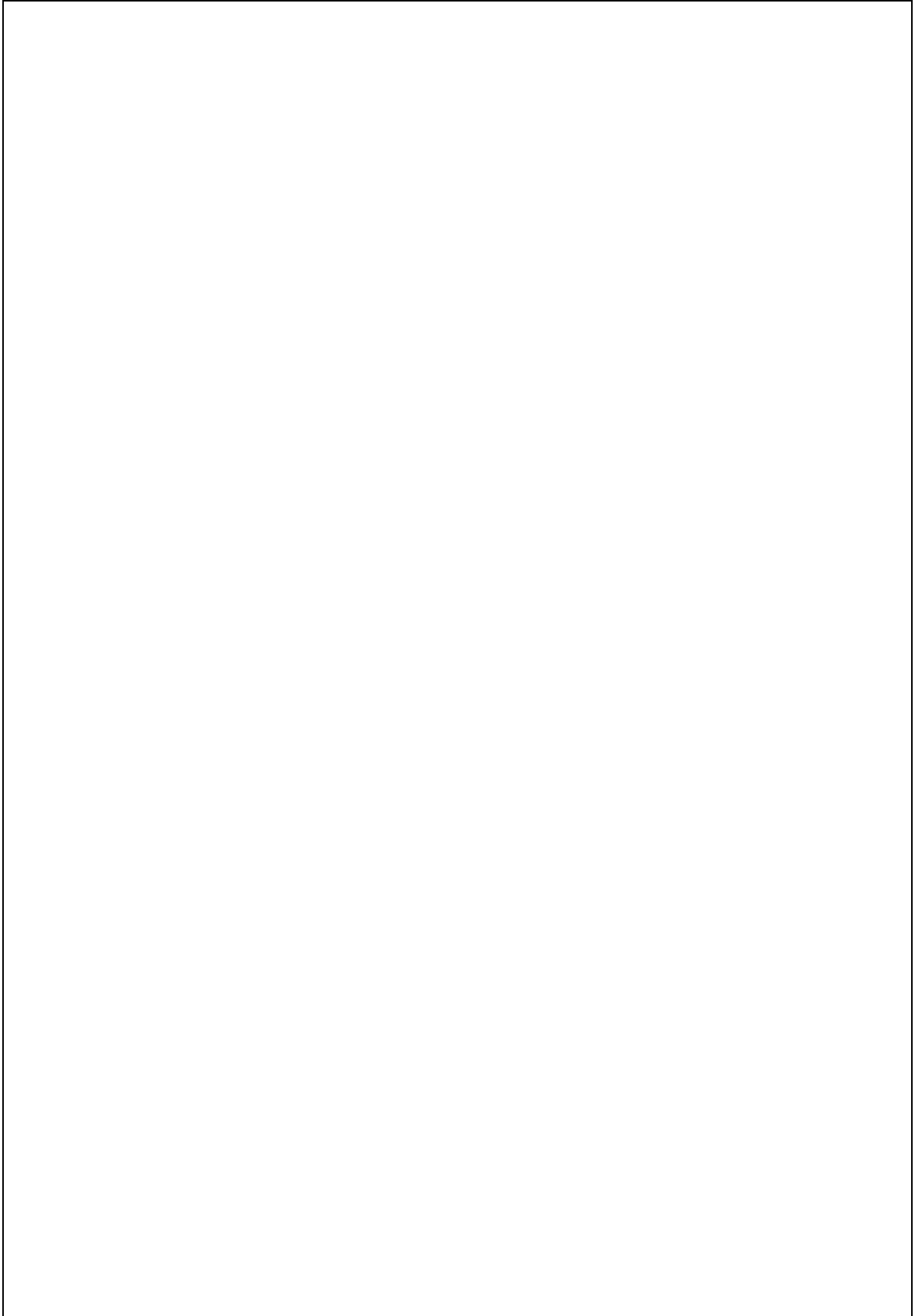
c) Time period: List years/cycles required.

**\*Note:** Only available data cycles should be listed. Please refer to: [Data available at the Research Data Centres](#).

d) Sample description: Identify the unit of analysis and population of interest.

e) Geography: Identify the level(s) of geography for which estimates will be produced (e.g., national, provincial, or lower).

f) Variable list: Provide a table listing expected variables to be used in the analysis (may be included in an appendix). Non-confidential codebooks may be available if needed; contact your local Statistics Canada Data Access analyst for more information.



g) Will datasets be merged or pooled together?

- If yes, please explain how the data from each source will be combined and utilized in this analysis. For example, you may pool data from the same source to increase your sample size or merge contextual data to the micro-records.
- Record linkages are not permitted in the Data Centres.
  - If you require a **new** microdata linkage for social data, do not submit your request through the MAP. Visit the [Social Data Linkage Environment](#) webpage for more information on social microdata linkage at Statistics Canada.
  - If you require a **new** microdata linkage for economic data, do not submit your request through the MAP. Visit the [Linkable File Environment \(LFE\)](#) webpage for more information on economic microdata linkage at Statistics Canada.



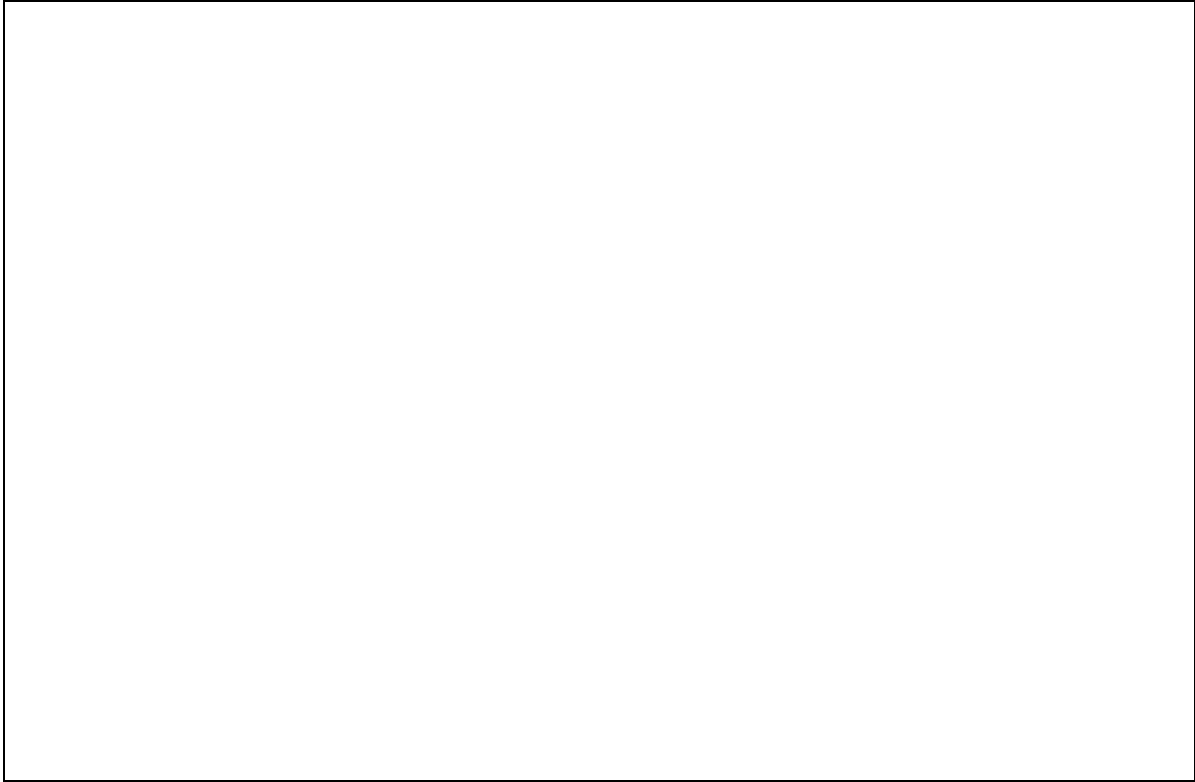
#### 4. Methodology:

a) Descriptive Statistics: Cross-tabulations and summary/descriptive tables and statistics must be:

- Kept to a minimum (justify the expected number of tables and lines);
- Support the analytical output and;
- Should be released at the end of the project, as much as possible.


*\*Note: Intermediary output is strongly discouraged as it may lead to residual disclosure issues.*

- List (a) the variables involved, (b) the number of dimensions in the tables, and (c) the number of tables.



b) Modelling:

- List (a) the dependent variable(s), independent variables, and (b) type of modelling being used (OLS, Logistic regression, etc.).



- c) Include a statement that applicable weights will be applied when using survey data.

- d) Is the project using Machine Learning techniques?

If yes: Describe in detail the type of machine learning techniques being used (e.g., supervised/non-supervised, neural networks, support vector machines (SVM) and tree-based methods including random forests, variable selection such as lasso, ridge, and elastic nets, etc.).

- e) Is the expected sample size sufficient to respect the confidentiality of the respondents?

**Describe requirements for any External Data:**

- Statistics Canada recognizes that in some cases, merging data from different sources has the potential to strengthen research by providing contextual information.  
*\*Note: Merging data does not mean linking records; see section 3g above.*
  - External Data refers to information that is collected by other organizations, departments of government and individuals for their own purposes.
  - Non-confidential Statistics Canada aggregate data tables (formerly CANSIM Tables) may be added to your research project with an approval.  
*\*Note: This does not apply to Public Use Microdata Files (PUMFs), which are not permitted*
- f) Please describe any external datasets that you plan to use, and how you will incorporate them into your analysis.

**5. Software requirements:** What software will be required?

- Not all software programs and packages are available for use in the SDAC/RDC.
- Approval of your proposal does not mean approval for software requested. All software packages and add-ons that are not currently approved for use are required to go through a separate approval process, which may take time.

- STATA, SPSS, SAS\* and R are available at all data centres.
  - **\*Please note:** SAS is currently available in SDAC but will be decommissioned in 2028, after which, it will not be available to SDAC researchers.
  - We encourage researchers to use open-source software such as R and Python.
- For a more exhaustive list of available software, please contact your local data center through [Contact information for the Data Access Division](#).



**6. Expected output:** Describe the products that will result from the proposed analysis (e.g., working paper, peer-reviewed journal article, book or chapter, thesis or dissertation, conference presentation, or commissioned report (e.g., government report)).

- If this project includes Indigenous data, please describe how the findings will be shared back with Indigenous communities and organizations.



**7. Expected start date and duration of the project**

**8. Source(s) of funding for this project:** Please list all sources of funding or indicate n/a if not applicable.

**9. Literature references**