

Statistics Canada Research Data Centres (RDCs)

Guide for Researchers under Agreement with Statistics Canada



October, 2005

PREFACE

This guide applies to researchers working under contract to the federal department Statistics Canada accessing data in the Research Data Centres (RDCs) located at universities throughout the country. The RDC Program intends for this guide to assist researchers working or who plan to work in an RDC as well as summarizing the role of the RDC staff, so that researchers understand all facets of the RDC.

The guide is organized as follows:

- [Chapter 1](#): Introduction to the Research Data Centres and the general application process for accessing the centres.
- [Chapter 2](#): Procedures to follow at each phase of the research process, starting from the project proposal to submission of the final product.
- [Chapter 3](#): Describes the procedures for protecting the confidentiality of respondents in statistical output. And, explains how to remove statistical outputs from the RDC.
- [Chapter 4](#): Provides website addresses if researcher requires additional information.
- [Appendices](#) providing more detailed information on some of the guidelines for disclosure analysis, proposal and research paper guidelines and contract information.

Any comments or questions can be directed to the RDC staff or you may contact:

Dr. Gustave Goldmann, RDC Program Manager Room 1710 Main Building Tunney's Pasture Statistics Canada Ottawa, ON K1A 0T6 Phone: (613) 951-1472 Fax: (613) 951-4942 Email: rdc-cdr@statcan.ca

TABLE OF CONTENTS

| | |
|--|------------------|
| PREFACE..... | 1 |
| CHAPTER 1 - INTRODUCTION TO THE RESEARCH DATA CENTRES AND APPLICATION PROCESS | 3 |
| <i>1.1 Background</i> | <i>3</i> |
| <i>1.2 Research Data Centres: Application Process</i> | <i>4</i> |
| 1.2.1 Research Proposal | 4 |
| 1.2.2 Approval of proposals based on four criteria | 5 |
| 1.2.3 Project Proposal Template..... | 5 |
| 1.2.4 Granting Access to an RDC for students..... | 5 |
| 1.2.5 Masters degree students..... | 6 |
| 1.2.6 PhD students..... | 6 |
| 1.2.7 Undergraduate Students | 6 |
| <i>1.3 Proposal Review Process</i> | <i>6</i> |
| 1.3.1 Committee Decision | 7 |
| 1.3.2 Clarification of Proposal is Requested by the Committee | 7 |
| 1.3.3 The Steps for Researchers to Take if the Strategic Programs and Joint Initiatives program officer requests clarification of your proposal..... | 7 |
| 1.3.4 What Happens if the Committee Members Request that You Revise and Resubmit your Proposal? | 7 |
| 1.3.5 Are there Circumstances where a Proposal has been Rejected?..... | 7 |
| <i>1.4 Proposal Approved.....</i> | <i>8</i> |
| 1.4.1 Security Clearance Procedures | 8 |
| 1.4.2 RDC Microdata Research Contract | 9 |
| 1.4.3 Orientation..... | 9 |
| 1.4.4 Oath (Oath of Office and Secrecy)..... | 9 |
| 1.4.5 Deemed Employee Status..... | 10 |
| CHAPTER 2 - THE LIFE CYCLE OF RESEARCH AT THE RDC..... | 11 |
| <i>2.1 Beginning your Project at the RDC</i> | <i>11</i> |
| 2.1.1 Physical Security of the Centre | 11 |
| 2.1.2 The Computer Facilities | 11 |
| <i>2.2 Guidelines for Statistical Analysis of Statistics Canada Surveys.....</i> | <i>12</i> |
| <i>2.3 Your Use of the RDC - Security.....</i> | <i>13</i> |
| 2.3.1 Getting Support | 13 |
| 2.3.2 Accessing the RDC | 13 |
| 2.3.3 System and Electronic Devices | 14 |
| 2.3.4 Data | 14 |
| <i>2.4 Your Use of the RDC – Procedures.....</i> | <i>15</i> |
| 2.4.1 Confidentiality and Release of Research Output..... | 15 |
| 2.4.2 Obtaining New Data..... | 16 |
| 2.4.3 New Research Project with Same Data | 16 |
| 2.4.4 Add or Remove a Team Member | 16 |
| 2.4.5 Amendment of the Contract Completion Date | 17 |
| 2.4.6 Contract Extensions..... | 17 |
| 2.4.7 Revision Contracts | 17 |
| 2.4.8 Transferring of Results..... | 18 |

| | | |
|--|---|-----------|
| 2.5 | <i>Completing the Contract</i> | 18 |
| 2.5.1 | Producing a Product for the Project..... | 18 |
| 2.6 | <i>What happens to the product(s) when submitted?</i> | 20 |
| 2.6.1 | Seminars and Conference Presentations..... | 21 |
| 2.6.2 | Related Papers and Research Reports | 21 |
| 2.7 | <i>Exit Process</i> | 21 |
| 2.8 | <i>Re-entering the RDC under a New Agreement</i> | 21 |
| CHAPTER 3 - DATA ACCESS, CONFIDENTIALITY, DISCLOSURE ANALYSIS POLICIES AND PROCEDURES | | 22 |
| 3.1 | <i>What is Disclosure?</i> | 22 |
| 3.1.1 | Types of Disclosure..... | 22 |
| 3.1.2 | Some Examples of Disclosure in Survey Data..... | 22 |
| 3.2 | <i>How Can Researchers Protect Against Disclosure?</i> | 23 |
| 3.3 | <i>General Risks for Confidentiality Problems</i> | 23 |
| 3.4 | <i>Disclosure Analysis</i> | 24 |
| 3.4.2 | Disclosure Analysis for Parametric Model Output..... | 24 |
| 3.4.3 | Disclosure for Non-parametric, Semi-parametric or Hierarchical Data | 25 |
| 3.4.4 | Disclosure Analysis for Variance-Covariance Matrices..... | 25 |
| 3.5 | <i>Use of Weights</i> | 25 |
| 3.6 | <i>Quality Guidelines for Releasing Data</i> | 26 |
| 3.7 | <i>Responsibilities</i> | 26 |
| 3.7.1 | The Researcher’s Responsibility | 26 |
| 3.7.2 | The Analyst’s Responsibility | 26 |
| 3.8 | <i>How to Remove Research Output from the RDC</i> | 26 |
| 3.8.1 | Disclosure Requests | 26 |
| 3.8.2 | Some Quick Tips for How Researchers Can Check for Disclosure | 27 |
| CHAPTER 4 - Acquiring More Information | | 28 |
| APPENDIX 1 - MORE ON DISCLOSURE AND DISCLOSURE RISK ... | | 29 |
| APPENDIX 2 - SAMPLE MICRODATA RESEARCH CONTRACT | | 30 |
| APPENDIX 4 - THE OATH | | 38 |
| APPENDIX 5 – SAMPLE RESEARCH CONTRACT AMENDMENT | | 39 |
| APPENDIX 6 - SAMPLE DISCLOSURE REQUEST FORM | | 41 |

CHAPTER 1 - INTRODUCTION TO THE RESEARCH DATA CENTRES AND APPLICATION PROCESS

1.1 Background

Decision-makers need an up-to-date and in-depth understanding of Canadian society to help them respond not only to today's needs, but to anticipate the needs of tomorrow as well. This need underlies the growing demand for analytical output from the rich sources of data collected by Statistics Canada.

The Research Data Centres (RDC) program is part of an initiative lead by Statistics Canada, the Social Sciences and Humanities Research Council (SSHRC) and a university consortia to help strengthen Canada's social research capacity and to support the policy research community.

RDCs provide researchers with approved access, in a secure university setting, to microdata from population and household surveys. The centres are staffed by Statistics Canada employees. They are operated under the provisions of the *Statistics Act* in accordance with all the confidentiality rules and are accessible only to researchers with approved projects who have been sworn in under the *Statistics Act* as 'deemed employees.'

RDCs are located throughout the country, so researchers do not need to travel to Ottawa to access Statistics Canada microdata.

Each Research Data Centre (RDC) employs an RDC Analyst who would be happy to advise you during the preparation of your proposal. The Analyst can assess your proposal to determine whether the required elements are present. The Analyst can also advise whether your project is appropriate for access to the detailed microdata. Also, the RDC Analyst can provide you with information about conducting research in the RDC environment and the data holdings of the RDC.

Please contact your [local RDC Analyst](#) for more information.

You may also find it helpful to access documentation for the surveys located on the Research Data Centre's website:

- <http://www.statcan.ca/english/rdc/whatdata.htm>
- or read about tips on proposal writing located at <http://www.statcan.ca/english/rdc/apply.htm> or <http://www.statcan.ca/english/rdc/apply.htm#guides>)

1.2 Research Data Centres: Application Process

Any researcher (or team) from academic organizations with legitimate research interests related to public policy development is eligible to apply to the RDC program. The detailed microdata kept by Statistics Canada in an RDC are protected under the [Statistics Act](#). Only employees or deemed employees of Statistics Canada may have access to these data. A process has been established to permit *academic researchers* with approved proposals to work with these data. Researchers must become deemed employees of Statistics Canada and produce a research paper that follows the Statistics Canada mandate (*Statistics Act sections 3, 19, 20, 22*).

To be granted access to the RDCs, a researcher must submit a research project proposal to SSHRC for peer and institutional review, undergo a security evaluation, and take an oath promising to protect confidentiality. The proposed project, once approved, will be the basis of a contract between the researcher and Statistics Canada.

1.2.1 Research Proposal

Research proposals are submitted to an adjudicating committee operating under the auspices of SSHRC and Statistics Canada. It is done electronically by accessing the submission form on the SSHRC website:

http://www.sshrc.ca/web/apply/application/rdc_application_e.asp

Researchers must submit a separate proposal for each new project that they wish to carry out in the RDC. In the case of significant changes in scope of the approved project, researchers may be asked to re-submit another proposal incorporating the changes.

To apply for RDC access, you also obtain [personal security clearance](#) from Statistics Canada (refer to the section [Security Clearance Procedures](#)).

SSHRC invites applications from individual researchers or research teams led by a principal applicant. The principal applicant is responsible for submitting application forms on behalf of the team.

Applicants must complete the Web-based CV (https://webapps.nserc.ca/sshrc/logon_ciss_e.htm) and application forms and include the attachments listed below. Before you begin your on-line application, please ensure that you have prepared all required attachments. These attachments include research contributions (maximum five pages) and the Project Proposal (maximum five pages).

To ensure a swift review of your proposal, applicants are encouraged to include all elements of the [Project Proposal Template](#).

1.2.2 Approval of proposals based on four criteria

- scientific merit and viability of the proposed research;
- viability of the methods to be applied— the data to be analyzed;
- demonstrated need for access to detailed microdata; and,
- expertise and ability of the researchers to carry out the work..

1.2.3 Project Proposal Template

The Project Proposal is a maximum of five pages and must include the following:

- Title of the Project
 - The project title
- Rationale and objectives of the study
 - Clearly identify the specific questions or objectives of the project
 - State how the research will contribute to the knowledge in the field of study
- Proposed data analysis and software requirements
 - What is the proposed statistical methodology? How is it suitable for this project?
 - What software will you use?
- Data Requirements
 - An explanation of why access to the confidential data (as opposed to public use microdata files) is necessary.
 - Which survey file/files or cycles are to be used?
 - Provide a statement that the confidential data file(s) identified is (are) in fact suitable for the proposed research.
 - What is the specific population of interest in the required data set(s)?
 - What are the variables to be used?
- Expected project start and end dates
 - Expected project start and end dates
- References
 - Sources used to cite your quotes used in the proposal or for specific analytical methods employed

1.2.4 Granting Access to an RDC for students

Access will be granted based on the nature of the proposal. For example if a proposal requires analysis of cross-sectional data sets this would not on its own be considered in depth enough to warrant access to RDC data.

In general, students listed as either principal or co-investigators on proposals (for thesis or dissertation research) should include in their application a letter of recommendation from their supervisor stating that they are capable of utilizing statistical software such as SPSS, STATA, or SAS and that the supervisor will be an active member of the research team supporting the student throughout the process.

Students are to send a copy in electronic format to lle@sshrc.ca, and a signed hard copy to:

Luc Lebrun, C/O SSHRC
RDC Program Officer
350 Albert Street
P.O. Box 1610
Ottawa, Ontario K1P 6G4

1.2.5 Masters degree students

Master degree students can apply as Principal Investigator (PI) or co-investigator, but their supervisor **must be part** of the research team and be included on the contract.

1.2.6 PhD students

All PhD students who request access to the RDC **must** include their Academic Advisor as the co-investigator on their research team.

1.2.7 Undergraduate Students

Access will only be granted to Honour's year undergraduate students. The Academic Supervisor/Advisor must be the Principal Investigator of the research team in order to grant access. The student then becomes a co-investigator (the PI does not necessarily have to access the RDC and can leave this data research solely to the student).

1.3 Proposal Review Process

The formal review process is coordinated by SSHRC. A three-member peer-review committee selected by SSHRC and Statistics Canada is formed for each proposal submitted. Two of the peers are from academic institutions with expertise in the field of the proposal. The third member is chosen from Statistics Canada.

The committee evaluates each proposal based on its scientific merit and recommends to Statistics Canada whether or not the project meets the established criteria (refer to the [criteria](#) outlined above). As such, the task of the committee differs fundamentally from other selection committees run by SSHRC. The committee does not select winners from a range of applicants, and there is no financial grant involved in this process. The committee acts as a gateway.

Research proposals that have undergone peer review from one of the tri-council funding agencies (SSHRC, CIHR, NSERC or FQRSC) will bypass peer review and move directly to institutional review.

1.3.1 Committee Decision

Within eight weeks of the date of application (sooner if possible), SSHRC will communicate the adjudication committee's decision to the project's principal applicant. The review process will produce one of two outcomes – acceptance or rejection; and the decision must be unanimous among all peers. The applicant is advised of the results via written correspondence. The decision rationale is included in the report of the evaluation results for the projects that are rejected. Researchers are free to address these concerns and resubmit the proposal.

1.3.2 Clarification of Proposal is Requested by the Committee

On occasion, committee members request clarification from the principal applicant if the proposal does not demonstrate, clearly, for example that their project requires access to detailed micro-data (i.e., that their analyses cannot be accomplished with publicly available micro-data). If clarification is requested of the principal applicant, then the Strategic Programs and Joint Initiatives program officer, will act as a liaison between the committee members and the principal applicant, and request the necessary information.

1.3.3 The Steps for Researchers to Take if the Strategic Programs and Joint Initiatives program officer requests clarification of your proposal

Provide the necessary clarification to the Strategic Programs and Joint Initiatives program officer as soon as possible so that the committee members can review your comments and continue to evaluate your proposal.

1.3.4 What Happens if the Committee Members Request that You Revise and Resubmit your Proposal?

Some proposals require substantial modifications and it is simpler to request that the principal applicant revise and resubmit their proposal in order to avoid many rounds of committee member evaluations. If the Strategic Programs and Joint Initiatives program officer contacts you for a revise and resubmit, you are invited to consult with the local RDC analyst if you are unclear about the revisions requested of you. Once you have modified your proposal, resubmit it online (http://www.sshrc.ca/web/apply/application/rdc_application_e.asp) to the SSHRC application page.

1.3.5 Are there Circumstances where a Proposal has been Rejected?

If a proposal is rejected by all committee members then the researcher is invited to resubmit their revised proposal. The **Strategic Programs and Joint Initiatives program officer**, Luc Lebrun, will inform the researcher that they can modify their proposal and re-submit through the proper SSHRC-channels. All rejected applications

must be officially re-submitted through the RDC electronic system and be subject to review by the overall evaluation committee. Researchers can expect that this process will take the same estimated time as their original proposal submission to SSHRC.

The only exceptions to the overall evaluation process apply to the CIHR, SSHRC or FQRSC-funded projects which need only be sent back to the **Strategic Programs and Joint Initiatives program officer**, Luc Lebrun, who will then forward the revised proposal directly to the relevant Statistics Canada Institutional reviewer for evaluation (omitting the Academic peer review step).

1.4 Proposal Approved

If your proposal is approved and you are given access to a Research Data Centre, your contract with Statistics Canada allows you to access only the microdata specified in your approved research project and only for the purpose of completing that project. You must submit a proposal for any subsequent research project that you wish to carry out at an RDC. In addition, SSHRC and Statistics Canada may ask for a new proposal if the scope of your research changes significantly.

Once the proposal is approved, researchers are required to get personal [security clearance](#) from Statistics Canada.

1.4.1 Security Clearance Procedures

Once a project is approved a number of security procedures must be followed:

- Statistics Canada will perform an Enhanced Reliability Check on any researcher who **needs to access its data**.
- Researchers will have to complete the [security clearance forms](#) within the presence of the Statistics Canada analyst at the RDC where the research will be conducted.
- The RDC analyst will send this form to Statistics Canada in Ottawa to be processed and will contact researchers to inform them of the results of the security check.
- The RDC analyst will invite the researcher, or group of researchers, for an orientation session to explain procedures at the RDC. (**All team members accessing the data**:
 1. require security clearance;
 2. must sign the contract;
 3. must attend an orientation session; and,
 4. must sign [The Oath or Affirmation of Office and Secrecy](#).

- During the orientation session, researchers will sign their contract with Statistics Canada and take the [Oath](#) or Affirmation of Office and Secrecy (refer to [Appendix 4](#)).

Note: Researchers with no active research contracts with Statistics Canada for more than a year are required to update their security clearance.

1.4.2 RDC Microdata Research Contract

The accepted proposal will become part of the contract between the researcher(s) and Statistics Canada. The contract specifies the following terms of access:

- Data sets to be provided by Statistics Canada (please note the contract grants researchers access only to the microdata specified in the approved research proposal).
- Project start and completion date.
- Agreement that the researchers abide by the RDC security and confidentiality requirements.
- In fulfillment of the contractual obligations, the researcher will provide a product to Statistics Canada once the contract end date has arrived.

1.4.3 Orientation

The orientation session is a presentation designed to familiarize researchers with the privilege and duty of becoming deemed employees of Statistics Canada and working in the RDC. It includes a discussion of the policies and procedures related to data confidentiality, disclosure analysis, and other information regarding the operation of the RDC. Procedures for obtaining the release of research output will also be discussed. Researchers will also sign the Oath of Office during the orientation.

The orientation session is not and should not be considered a substitute for a thorough reading of this document (Guide for Researchers).

1.4.4 Oath (Oath of Office and Secrecy)

Before accessing data, researchers are asked to take an Oath of Office (oath of secrecy) to become a deemed employee of Statistics Canada. This is the legal requirement to maintain data confidentiality. A copy of the oath can be found in [Appendix 4](#). As part of this process the researcher(s) agree(s) to:

“... not disclose or knowingly cause to be disclosed, by any means, any information obtained under the Statistics Act in such a manner that it is possible from the disclosure to relate the particulars obtained from any individual return to any identifiable individual person, business or organization” (Statistics Act).

This oath requires that researchers be personally accountable to uphold the confidentiality provisions of the Statistics Act, and not to reveal anything about individual respondents, either directly or indirectly. The oath is binding for life; hence even after researchers have completed their research contract, they may not reveal any confidential information. A violation of confidentiality, whether intentionally or accidentally, would put the RDC endeavour at risk, and reduce research opportunities. In this document, the majority of the discussion is intended to help prevent an accidental breach of confidentiality.

It is also important to remember that this oath is legally binding with repercussions if violated. If a researcher violates the confidentiality provisions of the Statistics Act, the researcher will lose all privileges at this and any other Statistics Canada RDC indeterminately. Other consequences and penalties are outlined in [Appendix 4](#). Note that Statistics Canada employees are subject to the same penalties when a violation of the oath occurs.

1.4.5 Deemed Employee Status

Once you have signed the Oath, you are a deemed employee of Statistics Canada for the duration of the microdata research contract. All deemed employees within the RDC Program are legally responsible for upholding the Statistics Act and all policies and procedures of Statistics Canada. Deemed employees are not permitted to take away with them any sensitive statistical information. They remain subject to the oath/affirmation of secrecy even after their project is terminated.

The contractual arrangements between Statistics Canada and deemed employees do not involve payment to the researcher but require:

- Statistics Canada is to provide access to confidential microdata (sensitive statistical information); and,
- The researcher is to deliver a product at the end of the contract.

CHAPTER 2 - THE LIFE CYCLE OF RESEARCH AT THE RDC

2.1 Beginning your Project at the RDC

All Statistics Canada's RDCs have particular elements in common that researchers should be aware of and adhere to. All RDCs are physically secure environments with an isolated computer network (i.e., no internet capabilities or other external capabilities). Statistics Canada staff, RDC analysts, are on site to assist researchers and to approve all output leaving the RDC. This chapter is designed to help researchers through each phase of the life cycle of a research project conducted in the RDC environment.

2.1.1 Physical Security of the Centre

The RDC is a physically secure facility to the extent that the researchers use it properly.

- Each researcher will be issued a security pass. One cannot gain access to the RDC without the pass.
- Do not share or lend your security pass to anyone else.
- Report lost or stolen security passes immediately to the RDC Analyst.
- No visitors. Researchers may not escort individuals who do not have security clearance into the centre. Only RDC Analysts can permit entry of any individual into the premise.
- Do not open the door for anyone (i.e. do not answer a knock/doorbell indicating that someone without a pass would like entry into the RDC).

2.1.2 The Computer Facilities

Researchers are not assigned specific workstations in the RDC. It is important to book a time in the RDC in advance to minimize conflicts over resources and workstations. As a courtesy to other users, please adhere to the booking as much as possible.

Contact the RDC Analyst to request the purchase and installation of new software or software upgrades. Researchers may not install their own copies of software onto any RDC workstation.

- Each computer in the RDC is a workstation to which researchers can log on and connect to the server.
- Data
- Do not operate laptop computers, PDAs, cell phones or any other storage devices in the lab area.
- Workstations do not have CD/floppy disk-write capabilities. Disk-writing can only be conducted by the RDC Analyst through secured terminals.
- It is strictly prohibited to connect any portable or mass storage device to a workstation.

2.2 Guidelines for Statistical Analysis of Statistics Canada Surveys

Statistics Canada household surveys, such as those whose data are available in the RDCs, are based on complex sample designs that include stratification, multiple stages of selection, and unequal probabilities of selection. Generally, commercially available statistical software packages ignore these complexities. (Some of these packages can make use of the weights, which do contain some design information, but do not compute correct variances.) However, software that accounts for the design complexities is available.

Ignoring the design complexities could have an impact on the results of the analysis of the data. As an example, because of special interest in certain subpopulations, a survey might sample some units at higher rates than other units; an analysis of the population as a whole that ignores these differential rates will generally produce biased results. As another example, clusters of units are often included in a sample, such as several people from the same household; this clustering may lead to correlation between the observations, which, if ignored, could lead to underestimation of the standard errors of parameter estimates.

Many standard analytical procedures have been adapted to incorporate the sample design information. However, there are some analytical procedures for which there is currently no recommended design-based approach. Research is ongoing into appropriate analytical methods for such cases. In the meantime, approaches developed for non-survey data would have to be used; however, for these cases, there still may be some facility for making use of the survey weights.

2.3 Your Use of the RDC - Security

2.3.1 Getting Support

RDC Analyst

The first line of support while working at the RDC comes from the RDC Analyst. Researchers can direct questions about the dataset to the RDC analyst. The analyst will either have the answer or will direct the question to other Statistics Canada employees. When managed properly, this type of knowledge exchange benefits both the researchers and the Statistics Canada data programs.

Data Analysis Resource Centre (DARC)

The second line of support is available through the Data Analysis Resource Centre (DARC) at Statistics Canada. There has been considerable research carried out by this group on the appropriateness of various statistical analyses for complex sample designed datasets produced by Statistics Canada. DARC can provide suggestions on suitable methods and software tools. DARC also conducts seminars and research on various methods specific to longitudinal data. These services are offered on a cost-recovery basis. For more details, please contact your RDC analyst.

2.3.2 Accessing the RDC

As a Statistics Canada's deemed employee, it is your responsibility (as well as the RDC analyst's) to maintain the RDCs' security:

- Breaches in security or lost door keys should be reported to the RDC analyst immediately.
- RDC entrance information must not be shared.
- Consult with the RDC analyst for the sign-in procedures for the RDC.
- No visitors are permitted into the secure RDC facilities.

2.3.3 System and Electronic Devices

- Do not operate any personal electronic device such as laptop computers, palm pilots, cellular phones or other devices with optical beam capability inside the RDC secure premises. In case of emergency, cellular phones, or similar device may be used inside the RDC but away from the workstation.
- You are accountable for your working files and how you use the system. With this accountability comes the responsibility of protecting your log-on information.
- Make sure to log off after every work session. If you need to be away from your workstation during the session, all work should be saved and the computer locked to prevent others from accessing it.
- Save all work in your assigned electronic folder. This aides in the disclosure avoidance that the RDC analyst must do before your output can be released, project folder backup and system maintenance.

2.3.4 Data

- Do not discuss confidential information at any time outside the RDC, whether over the telephone, by e-mail (or other electronic media) or in person.
- Do not ask other researchers at the RDC for access to their data sets. Researchers are granted access only to the data for which access has been approved. Written requests or a new proposal is required for access to other data sets.
- Do not give other researchers, other than members of your team, access to any of your secondary datasets or output files. See the RDC Analyst for ways to share files with team members electronically without sharing passwords.
- Do not carry out analyses for colleagues or other researchers who do not have approved RDC research projects or who are not deemed employees.
- Do not use RDC data files to conduct data analysis outside the mandate of the approved project.
- Please place any output containing confidential data that you wish to discard into a designated receptacle, shredder, or return to the RDC Analyst.

2.4 Your Use of the RDC – Procedures

2.4.1 Confidentiality and Release of Research Output

Statistics Canada considers it important not only to avoid disclosure of confidential information, but also to avoid the *perception* of disclosure. It is the trust of the respondents that makes it possible for Statistics Canada to provide valuable data on the socio-economic condition of Canadian society.

The following is a brief summary of Statistics Canada policies regarding confidentiality and obtaining release of the research output.

- *Absolutely no detailed microdata may leave the facilities.*
- *Absolutely no research results may leave the facilities without being examined and approved by the RDC analyst.*
- *All output leaving the RDC must undergo disclosure analysis, first by the researcher and then by the RDC Analyst. Disclosure analysis is the means by which researchers ensure that any material removed from the RDC does not pose a disclosure risk of data confidentiality. See [Chapter 3](#), [Appendix 1](#) and [Appendix 6](#) for a complete discussion of the topic.*

Researchers must complete a [Disclosure Request Form](#) each time they make a disclosure request. Researchers must provide description of variables used, new variables created, documentation of data sets and programs used in producing output to help the RDC Analyst better understand the materials requested for release. Working closely with the RDC Analyst during the disclosure analysis process will help to avoid misunderstandings and speed up the final approval for release.

Disclosure of individual cases in any manner – in research papers, via email, phone or fax, or in casual conversation – with persons outside designated facilities is prohibited. This includes researchers working on other projects who have deemed employee status.

Research output that emphasizes model output rather than tabular (descriptive) output are encouraged in order to reduce disclosure risk.

Removal of “intermediary output” from the RDC is discouraged. Intermediary output is typically produced during the exploratory and model development stages of the analysis. These outputs often consist of detailed tables of descriptive statistics. Such outputs often present more detailed information than final model outputs; their removal can greatly increase the risk of disclosure, especially residual disclosure.

Within the secure premises of the RDC, output of any type can be printed, but it must not be removed from the RDC before being examined and approved by the RDC analyst. Output without disclosure control will be printed on coloured paper (usually green) to avoid accidental confusion with outside material brought into the centre or output already approved to leave the centre. Coloured paper (usually green) **DOES NOT LEAVE THE RDC**. At the end of a

session, materials should be returned to the RDC analysts for secure storage until the next visit. This material will be shredded by the researcher or when the project is completed.

Researchers must comply with Statistics Canada rules and regulations on data confidentiality when removing output from the RDC.

2.4.2 Obtaining New Data

On rare occasions, it may be appropriate to access another data set to complete the project. An appropriate case may be that the sample size in the originally selected data set is not adequate but another data set contains similar questions and more respondents. The researcher may also want to verify their findings from one data set using another dataset. The nature of the project or objective of the research questions must not change significantly.

If the researcher wishes to add a data set to their project under these conditions, they must submit a written request to the RDC Analyst detailing the rationale for accessing a new data set and a commitment to maintain the objectives of the original proposal.

Provide a detailed discussion when requesting new data be added to your project. This request will be subject to the formal project review process. If the proposed use of the new data does not fit within the original proposal as judged by Statistics Canada, you will be asked to begin a new project by submitting a new proposal.

The contract does not have to be amended to add a new cycle of a longitudinal data set, only for a new data set or if there are changes to the research questions.

2.4.3 New Research Project with Same Data

While working on a project a researcher can begin to develop new ideas for other analyses. If these analyses do not fit under the original objective of the initial proposal a new proposal will be required. Do not start a new project without submitting a proposal to the review process. Violation of this rule will result in immediate withdrawal of data access privileges including data access to projects currently under contract.

2.4.4 Add or Remove a Team Member

After approval of the research proposal, researchers may make changes to the research team. To remove a research team member from the project, please provide a written memorandum to the RDC analyst. Both the principal investigator and the team member being removed from the contract need to sign the amendment. To add a new researcher, please provide a written memorandum along with a curriculum vitae of the researcher to the RDC Analyst. The new member will be required to complete the **PERSONNEL SCREENING, CONSENT AND AUTHORIZATION FORM**; part of the process for the researcher to undergo a Reliability Check.

And they will also need to take the oath/affirmation of secrecy and attend an RDC Program orientation session.

2.4.5 Amendment of the Contract Completion Date

There are two ways to grant RDC data access after contract expiry: by contract extension or by a revision contract.

2.4.6 Contract Extensions

This type of extension applies when the principal investigator realizes that they have significantly underestimated the time needed to complete the project or other factors may be preventing them from completing their project by the date specified in the contract.

Contract extensions occur when a project reaches its contract end date but the research team has not yet completed enough analysis to prepare their product. There are valid reasons for a project extension including:

- The data analysis is in the final stages and the extension will cover the completion of the data analysis;
- Funding or staffing shortages have prevented progress on the contract to date (including leaves of absence, long-term illnesses, etc.);
- There have been complications in obtaining data to complete the analysis.

The contract can be extended for up to one year past the contract due date at the request of the principal investigator.

Anticipate the need for an extension of your project and provide your RDC analyst with a written justification well in advance of the contract expiring.

2.4.7 Revision Contracts

A revision contract for data access may be obtained only in the following two cases:

- (1) A researcher has submitted to a scholarly journal a manuscript based on results obtained from their RDC research;
 - o The researcher has received a “revise and resubmit” from the editors of that journal *that requires them to return to the data*; and,
 - o The researcher’s original project contract *has expired*.
- (2) A student has submitted a dissertation or thesis to their academic review committee;

- o The student has received comments, usually through marked up copies of the text, and must now submit a written request to the RDC analyst for permission to re-enter the RDC in order to perform additional work with the data.
- o The student's original project contract *has expired*.

Access under a Revision Contract is provided with the understanding that the research team is undertaking only the work suggested by the referees and not undertaking a new research project. They are generally used when a project has expired and their manuscript has already been submitted to Statistics Canada in fulfilment of their contractual obligations.

- Revision contracts usually last up to six months from start to finish. A product is required at the end of the contract even if the content is very similar to their original manuscript.
- In addition to providing the information required on the *Contract Extension/ Revision Contract Request Form*, the Principal Investigator must provide a copy of the referee reports as evidence that the project continues to have scientific merit and requires re-access to the microdata. If you are uncomfortable sending these reports in their entirety, a summary of these comments will suffice.

2.4.8 Transferring of Results

Results may be transferred to another RDC without disclosure analysis when used by other team members in another RDC. Transfers require that the results be encrypted by the RDC analyst to encrypt the results before they are recorded on the transportation medium (CD or DVD). Please schedule time for transferring files with the RDC Analyst.

2.5 Completing the Contract

RDC Program at Statistics Canada has a mandate to provide a publicly available document from all research conducted on its data sets and this requirement extends to research conducted by deemed employees in exchange for data access. Researchers are responsible for producing a product from their research. See the contract, [Appendix 3](#), section (5), "Limitations on uses of the microdata file and proposed output".

Approximately one to two months prior to contract expiry date a reminder letter (signed by the Program Manager) will be sent by email to the Principal Investigator.

2.5.1 Producing a Product for the Project

Researchers with approved projects in the RDCs are granted access to the detailed micro data as deemed employees of Statistics Canada under the terms of the Statistics Act. The Act stipulates that only employees and deemed employees of Statistics Canada may have access to the detailed

(non-public use) micro data. To be a deemed employee of Statistics Canada implies that some product or service must be rendered to the agency. The product is defined in the research proposal that was submitted as part of the application process and it is stipulated in the contract that is entered into between the researcher(s) and Statistics Canada.

A project is complete¹ when the Principal Investigator submits the product to Statistics Canada thus fulfilling their contractual obligations. The Research Data Centre Program has redefined its policy on the type(s) of products they would like researchers to submit in fulfilment of their contractual obligations for data access. For the most part, it means that instead of submitting an RDC working paper, researchers now have the option to submit a number of different products such as a journal article as their final product for the contract. Researchers are always encouraged to submit all final products to us so that we track these as measures of success for the RDC program.

There are five types of products that a researcher can produce as a product for their project.

1. RDC working paper
2. Peer-reviewed journal article
3. Book or book chapter
4. Graduate level thesis or dissertation²
5. Commissioned report (e.g., a government commissioned report)

If a researcher identifies an alternate possibility, this should be discussed with an RDC Analyst supervisor before the contract is written. Changes to contract wording of the acceptable product will not be made retroactively, only proactively for new contracts.

If a researcher would like to submit a product other than a Statistics Canada working paper as the main product from their already existing contract, they should discuss this with their RDC Analyst who will be submitting the product to the RDC Manuscript Coordinator. .

¹ **Note:** the contract end date does not always coincide with the project completion date. The contract end date dictates when data access must stop. The researchers may continue to work on the project outside the RDC environment to complete the writing phase. A project ends when the main product is submitted.

² If a member of a research team wishes to use the analysis for a *thesis or dissertation*, a separate contract should be written to cover their analysis, even if it was not indicated separately on the original proposal. The student should submit a one-page description of the analysis plan explaining how their work is consistent with the original group proposal and sign a contract along with their supervisor.

2.6 What happens to the product(s) when submitted?

This is an important new development: Not all products will be peer and institutionally reviewed.³ Instead, the product will be given to a Triage committee. The Triage committee will identify the ‘gems’ that we would like to showcase through a Statistics Canada flagship publication venue. What is a gem? It is a product that has wide appeal to a Canadian audience with a clear message about social trends or policy. There are three outcomes from the Triage process:

1. The product is not appropriate for a Statistics Canada flagship publication, for one reason or another. For example, the results of the analysis did not identify any conclusions or the analysis was not feasible given the data. The Principal Investigator will be notified by email from Statistics Canada head office that the product was received and that their contractual obligations are fulfilled. The local RDC Analyst will receive a carbon copy of this message.
2. The product is a good product but it is not of general interest to a Canadian audience. The Principal Investigator will be notified by email from Statistics Canada head office that the product was received and that their contractual obligations are fulfilled. This product will be included in a bibliography of research conducted in the RDC. The local RDC Analyst will receive a carbon copy of this message.
3. The product is a ‘gem’ and Statistics Canada is interested in publishing it in a flagship publication. The Principal Investigator will be notified by written letter from Statistics Canada head office that the product was received and that their contractual obligations are fulfilled. That letter will also invite to the Principal Investigator to consider a Statistics Canada publication under one of three possible scenarios dependent upon their other intentions for publication. The three publication scenarios are:
 - (a) A joint release between a peer-reviewed journal and the Statistics Canada Daily. On the day the journal publishes the article, an advertisement for this will be listed in the Daily along with a brief description of the research.
 - (b) A release in the Statistics Canada Daily after a journal publication has occurred.
 - (c) A publication in a Statistics Canada flagship publication such as Canadian Social Trends, Health Reports, Perspectives on Labour and Income. The timing of this publication can be postponed to accommodate a peer-review journal publication first if requested by the Principal Investigator.

If the Principal Investigator is interested in pursuing one of these publication options and re-access to the data is required, either an [extension](#) or a [revision contract](#) will be prepared depending on the circumstances.

³ Some researchers who are new to the data set utilized in their project may have concerns about how they have analyzed the data. The Principal Investigator has the right to request an institutional review of their paper at any time after the submission of the final product to Statistics Canada. If a researcher would like an institutional review, the RDC Analyst should notify the RDC Manuscript Coordinator who will make the necessary arrangements.

2.6.1 Seminars and Conference Presentations

Results may be presented at seminars or conferences. All work presented must reflect the proposal. And, all results must undergo disclosure analysis by the RDC Analyst.

Please give the RDC Analyst a copy of all presentations so an inventory can be maintained.

2.6.2 Related Papers and Research Reports

Beyond the RDC product produced, you are requested to provide a reference, a copy of the abstract, or the full document of any subsequent papers produced from the original RDC product and analysis. Statistics Canada hopes to maintain a bibliography of publications resulting from the RDC program. Papers, abstracts or references can be sent to the RDC analyst or to the Program Manager.

2.7 Exit Process

When submitting the final paper, researchers are asked to provide secondary data sets, model programs, relevant documentation, output, etc used in creating the product. This information will be archived and can be accessed during the [revision contract](#).

At the time the project is complete, the Academic Director of your RDC will be interested in posting your product (or a reference to it) on the local RDC website. This is negotiable and is done on a case by case basis.

See your RDC Analyst about cleaning out your storage area, archiving files and returning your security card and/ or locker keys.

2.8 Re-entering the RDC under a New Agreement

To access the RDC to begin a second or subsequent research project, researchers will follow the [application process](#) outlined earlier in this document by submitting a research proposal through the Social Science and Humanities Research Council (SSHRC) web site.

After the proposal is approved, the research team members will sign a new contract and may be asked to review the orientation material with the RDC Analyst and reaffirm the oath before accessing the data.

CHAPTER 3 - DATA ACCESS, CONFIDENTIALITY, DISCLOSURE ANALYSIS POLICIES AND PROCEDURES

The material in this chapter applies only to output that you intend to remove from the RDC. Any amount of output can be produced and stored (up to available capacity) for use on-site. This chapter presents policies and procedures to follow to clear research results for release.

[Appendix 1](#) provides examples of disclosure problems and suggested approaches in correcting them. The goal of disclosure avoidance is to protect the information provided by respondents while presenting the least possible hindrance to research pursuits. Ultimately, users of confidential micro data files are responsible for avoiding disclosure of confidential data and *minimizing the risk of disclosure*. The RDC Analyst and researchers will work together to find solutions to data confidentiality problems.

3.1 What is Disclosure?

Disclosure occurs when data that can be attributed to individual respondents (e.g., persons, households, businesses, other organizations), are released.

3.1.1 Types of Disclosure

There are three types of disclosure; Identity, Attribute and Residual.

Identity disclosure occurs when an individual can be identified from the released output, leading to information being provided about that identified subject.

Attribute disclosure occurs when confidential information is revealed and can be attributed to an individual. It is not necessary for a specific individual to be identified or for a specific value to be given for attribute disclosure to occur. For example, publishing a narrow range for the salary of persons exercising a particular profession in one region may constitute a disclosure.

Residual disclosure can occur when released information can be combined to obtain confidential data.). Care must be taken to examine all output to be released. While a table on its own might not disclose confidential information, disclosure can occur by combining information from several sources, including external ones. (e.g., suppressed data in one table can be derived from other tables).

3.1.2 Some Examples of Disclosure in Survey Data

- A well-known personality, e.g., a professional athlete, is selected in a survey and;
- Information published about her community, such as the highest reported income in that community, almost certainly was reported by her. (Identity disclosure.)

- Results from a longitudinal survey highlight one household with a highly unusual migration pattern, leading to its identification. (Identity disclosure.)
- The parents of a 16-year-old selected in the sample see a table showing that all sampled 16-year-old respondents in their region have tried drugs. (Attribute disclosure.)
- A newspaper article relates a 37-year-old widower's complaints about being surveyed, and there are only two sampled 30 to 39-year-old widowers in survey cross-tabulations. (Eventually leading to identity and/or attribute disclosure.)
- By combining several results a person identifies information that was purposely excluded from the Public Use Microdata File (PUMF) because it presented too high a disclosure risk (e.g., the country of birth of recent immigrants).

Note that even the appearance of disclosure can tarnish a statistical organization's reputation with respect to confidentiality. Damage could occur even if it turned out that the wrong person or household had been identified in the first two examples. Refuting a mistaken identification may increase the risk of exposing the real respondents.

3.2 How Can Researchers Protect Against Disclosure?

Disclosure analysis is the term applied to the examination of output researchers would like to remove from the RDC site. It involves a careful look not only at whether obvious identification of individual cases has occurred, but whether information about individual cases can be inferred or deduced from the output. This chapter outlines some broad approaches to disclosure analysis that researchers can use to prepare output for release from the RDC. [Appendix 1](#) presents some examples of specific problems and approaches to solve them.

The most important policy is:

Absolutely no data and/or research results may leave the facilities without being examined and approved by the RDC Analyst.

3.3 General Risks for Confidentiality Problems

Researchers should keep in mind that data confidentiality is primarily a problem for descriptive data, tables of magnitude, individual statistics, etc. It tends not to be a problem for multivariate analysis results (as regression coefficients) except for a few special cases. Confidentiality is also a problem when you are using small geographical identifiers or narrowly defined population groups.

3.4 Disclosure Analysis

3.4.1 Disclosure Analysis for Tabular Output

Tabular output is what is often called ‘descriptive statistics.’ Researchers should limit the requests for disclosure clearance of tabular output to the minimum necessary to describe the sample used in models, and how it might compare to an underlying population.

Tables of summary statistics (means, variances, and related statistics) intended to describe the distributions of variables present certain special issues:

- Tables should not contain low frequency cells. Unless otherwise specified in the survey documentation, this means less than 5 observations in a cell.
- Take care with “full cells” (In two-way tables these occur when all the respondents in a particular row or column are concentrated in a single cell) or “zero cells”. Cells with no respondents can pose a particular problem with sensitive variables. These are variables such as income or health problems not generally known by the public.
- For many quantitative variables the maximum and minimum values must not be released. This applies particularly to sensitive variables related to income or consumption. Maximum values may also present a risk for apparently harmless variables, such as household size or age, as extreme values can lead to identity disclosure.
- Care must be taken in presenting other statistics intended to describe the shape of distributions. Where possible, present general statistics on distributional shape (e.g., skewness, kurtosis measures) rather than showing anything closely related to individual observations (e.g., the 99th percentile if small sample). Tables or graphs that report quintiles of distributions should be discussed with the RDC analyst.

If the researcher or the RDC Analyst finds a disclosure risk, modification of the output will be necessary (e.g., collapsing categories). This will avoid disclosure risk at the expense of output detail, and is the preferred course of action.

3.4.2 Disclosure Analysis for Parametric Model Output

Parameter estimates from some models, or the accompanying diagnostics generated by analysis packages, provide sufficient information to enable the re-creation of simpler statistics such as means and totals for cross-classes of variables. This is particularly true of so-called “saturated models” using binary coded variables, i.e., models that include all main effects and possible interaction terms. If the models or accompanying diagnostics provide much detail, it may be necessary to produce the underlying (un-weighted) summary results or tables for purposes of disclosure analysis.

If the researcher or the RDC analyst finds disclosure risk, modification of the output and possibly the model specification may be required. For example, researchers may be asked not to report the values of the affected coefficients (a good example would be the estimates of the fixed-effects in a fixed effects model). Or researchers may report that the coefficient is in a certain range (e.g., positive and significant). The RDC analyst and the researcher will work together to ensure, as much as possible, that the research results remain meaningful after any such modifications.

3.4.3 Disclosure for Non-parametric, Semi-parametric or Hierarchical Data

Non-parametric or semi-parametric estimates (for example, those used in survival analysis or kernel estimators used in discrete choice models) may pose special disclosure risks. An example is the now increasingly popular hierarchical linear model. This is likely to happen if the clusters contain small numbers of observations. Special care must be taken with output such as these. The RDC analyst and researcher will work together to avoid disclosure risk in these cases.

3.4.4 Disclosure Analysis for Variance-Covariance Matrices

While Statistics Canada prefers that all data analysis be completed in the RDC, some researchers will prefer to leave the RDC with variance-covariance matrices in order to conduct statistical inference and hypothesis testing. This is permitted, provided that the variance-covariance matrix adheres to the data confidentiality guidelines for tabular output and tables of magnitude. See [Appendix 1](#) for guidelines.

3.5 Use of Weights

Statistics Canada household surveys like those found in the RDC, are based upon complex sample designs with stratification, multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used. In order for survey estimates to minimize bias, the survey weights must be used.

Weighting brings results from a sample to the level of the population. Even when population estimates are not the main interest weighting corrects for sample non-representativity arising from the survey design (as over-sampling population sub-groups) and from data collection and processing operations (e.g., higher non-response among young adults). Weighting also assists in the protection of confidentiality because one case may no longer represent one respondent. In very limited circumstances un-weighted results can be released, such as analytic procedures that can not assimilate the use of weights. Researchers have to provide a written justification for requesting un-weighted results to be released.

It is important to include a statement in research manuscripts and presentations informing your audience that estimates may be influenced by the survey design and that population inferences are not valid when based on un-weighted results.

Please consult the documentation accompanying the survey for more information on the use of weights.

3.6 Quality Guidelines for Releasing Data

While data confidentiality is closely related to data quality, it is the responsibility of the researcher as RDC analysts do not screen output for data quality.

Documentation for each survey contains guidelines on what the minimum cell size should be to obtain acceptable estimates and maintain data quality. Before releasing and/or publishing any estimate, users should first determine the quality levels of the data set they are using. **The quality levels are acceptable, marginal and unacceptable.** All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by warning to caution subsequent users.

Please consult the documentation accompanying the survey for more information on data quality.

3.7 Responsibilities

3.7.1 The Researcher's Responsibility

Researchers are responsible for applying the rules and regulations for disclosure analysis as specified by Statistics Canada. Limiting the number of requests will reduce the risk of disclosure and facilitate the timeliness of processing. Disclosure analysis involves a certain amount of processing time, researchers should consider this when planning their activities and submitting output for release. A three day turn around time is typical; however, this time can vary depending on staffing resources and the number of other submitted disclosure requests.

3.7.2 The Analyst's Responsibility

A major responsibility of the RDC Analyst is to control the output released and ensure that confidentiality is not breached. When a problem of potential disclosure of confidential information is discovered in the results presented for review, RDC Analysts will do their best to find a solution. All attempts will be made in collaboration with the researchers. But overall, the RDC Program must maintain confidentiality and sometimes no release may be the outcome.

3.8 How to Remove Research Output from the RDC

You will be asked to provide materials that allow Statistics Canada to document all approved releases of research output. The documentation must indicate that the output was produced as part of an approved project (and which project), and it must allow Statistics Canada to document that published output does not disclose confidential information.

3.8.1 Disclosure Requests

At the beginning of your project you will be assigned a project computer directory on the network where all data sets, programming syntax, logs, and output should be stored. The RDC

Analyst will provide you with two folders ('to be vetted' and 'vetted'⁴) in your project directory. The 'to be vetted' folder is where you will copy the output for disclosure by the RDC Analyst as well as the Disclosure Request Form. Once the output has been vetted a copy will be placed in the 'vetted' folder and e-mailed to the researcher.

When you have output to be removed from the RDC:

1. Make an appointment with the RDC analyst to discuss the clearance request.
2. Fill out a Disclosure Request Form (see [Appendix 6](#) for a copy of the form). You are required to provide:
 - The survey name, weighting information, types of methods and sub-sample characteristics (or location of the files)
 - Supporting documentation files (or location of the files)
 - Information about possible residual disclosure

Please properly document the output. This can be done by ensuring variables are clearly labelled and cell sizes below the minimum requirement are collapsed or re-categorized. The researcher is to ensure that mistakes in the output have been corrected and specific disclosure issues for the survey are addressed.

Please be prepared to discuss the clearance request with the analyst performing disclosure analysis. This may require changes in the output to satisfy clearance requirements. Working interactively with the analyst performing disclosure analysis will maximize learning on both sides, avoid costly misunderstandings, and speed the review process. The clearer and more complete the request is, the sooner it can be released.

3.8.2 Some Quick Tips for How Researchers Can Check for Disclosure

Verifying these elements can speed up the vetting process:

- Provide un-weighted frequency counts (in an identical separate file).
- Do not submit tables with small cell sizes.
- Restrict cross-tabular analysis to two or three dimensions.
- Do not submit listings of cases or show graphs with outliers (i.e. no minimum or maximums).
- Provide supporting information for the correlation matrices.
- Be cautious when using small subgroups or small areas.
- Limit the release of tabular output before the end of your project.
- Schedule time with an RDC analyst well in advance.
- **Check documentation for survey specific rules.**

Refer to [Appendix 1](#) for further instructions on approaching disclosure problems.

⁴ The names of these folders may vary between centres.

CHAPTER 4 - Acquiring More Information

If more information is required about the RDC Network please log on to:
<http://www.statcan.gc.ca/english/rdc/network.htm>

If more information is required about the RDC Program please log on to:
<http://www.statcan.gc.ca/english/rdc/index.htm>

If more information is required about the Statistic Canada's data sets or research tools please log on to:
<http://www.statcan.gc.ca/english/rdc/whatdata.htm>

For an up-to-date list of the projects currently being conducted at the various centres or for abstracts and citation details of research papers published by researchers involved in the program, please log on to:
<http://www.statcan.gc.ca/english/rdc/rdcprojects.htm>

We are pleased to welcome you to the RDC program and facilities. We hope that your experience with the Program is going to be very profitable and rewarding for you and your research team.

APPENDIX 1 - MORE ON DISCLOSURE AND DISCLOSURE RISK

All variables on a database can be categorized according to their importance to data confidentiality:

Direct identifiers: Variables such as name, address or telephone number, etc. that provide an explicit link to a respondent. These are all stripped from the master files to which you will have access.

Indirect identifiers: Variables such as age, sex, marital status, area of residence, occupation, type of business etc. that could be used to identify an individual.

Sensitive variables: Characteristics relating to respondents' private lives, or business, which are not usually known.

These variables seem harmless on their own but used together could reveal information about individuals. For example, consider the case of drug experimentation during adolescence. The parents of a 16 year-old respondent may see a table showing that all sampled 16 year olds (indirect identifier) in their socio-economic group (indirect identifier) have experimented with drugs (sensitive variable). They thus know that their 16-year-old has experimented with drugs.

Data confidentiality is primarily a problem for frequency data and tables. It tends not to be a problem for correlative or causal analysis results such as regression coefficients. But it can occur in any kind of data output.

General rules to apply at all times:

- Use weighted data: the disclosure risk is reduced when weights are used to generate output.
- Do not report statistics or cells with fewer than five respondents (or the number reported in the documentation of the survey you are using, whichever is less).
- No anecdotal information may be given about specific respondents.

APPENDIX 2 - SAMPLE MICRODATA RESEARCH CONTRACT

Project ID number:

Revised, March 16, 2005

MICRODATA RESEARCH CONTRACT

BETWEEN:

HER MAJESTY THE QUEEN IN RIGHT OF CANADA,
represented by the Minister of Industry, designated as the Minister
for purposes of the Statistics Act, (hereinafter referred to as
"Statistics Canada"),

AND:

(Name and affiliation of Principal Investigator),
(hereinafter referred to as the "Principal Investigator"),

and,

(Name and affiliation of Co-investigators)
(hereinafter referred to as the "Co-investigators"),

WHEREAS Statistics Canada requires the services of the Principal Investigator to undertake statistical research and analysis on *(Name of the Microdata File)* to fulfill its mandate under the *Statistics Act*;

AND WHEREAS to perform these services and to have access to confidential information, the Principal Investigator and the Co-investigators must become "Deemed Employees" of Statistics Canada and are required to take the Oath of Secrecy;

AND WHEREAS Statistics Canada wishes to make clear the terms and conditions under which access to the microdata will be granted;

NOW THEREFORE the Parties agree as follows:

SERVICE PROVIDED BY PRINCIPAL INVESTIGATOR

1. (1) The Principal Investigator will carry out the research project set out in Appendix "A" and provide the report described under "Proposed Output".

- (2) It is understood that this is a contract for the performance of a service and the Principal Investigator and Co-investigators are engaged for the sole purpose of providing that service.

CONDITIONS OF ACCESS TO THE MICRODATA

2. The Principal Investigator and the Co-investigators must undergo an enhanced reliability check satisfactory to Statistics Canada and take the oath of secrecy in order to obtain access to the non identifiable microdata file required to perform the analysis pursuant to this contract.
3. (1) Access to the non identifiable microdata file (no names, addresses or identifying numbers) and associated documentation shall be provided on Statistics Canada premises, which includes Headquarters and the Statistics Canada Regional Offices during normal hours of operation, Monday to Friday, and the Research Data Centres.

(2) The Principal Investigator and Co-Investigators will be provided with the necessary computing facilities, software and documentation as is reasonably necessary to complete the research and analysis pursuant to this contract.

DEPARTMENTAL REPRESENTATIVE

4. The Director of the (*name of the division*) or the Manager of the Research Data Centre Program is the designated Statistics Canada representative responsible for the administration of this contract.

LIMITATIONS ON USES OF THE MICRODATA FILE AND PROPOSED OUTPUT

5. (1) The Principal Investigator and the Co-investigators shall not use or disclose any of the information obtained or produced pursuant to this contract for any administrative or regulatory purposes.

(2) Access to the microdata file is being provided for the statistical and research purpose outlined in the proposal attached as Appendix A and the microdata file shall not be used for any other purposes without the prior written consent of Statistics Canada.

(3) The Principal Investigator and the Co-investigators shall not disclose any of the information from the individual records obtained or produced

pursuant to this contract to anyone other than current Statistics Canada employees.

- (4) The Principal Investigator and the Co-investigators shall ensure that no attempts are made to link the microdata file to any other files in order to relate the particulars to any identifiable individual person, business or organization.
- (5) The Proposed Output must meet the requirements of both peer and institutional review prior to being released by Statistics Canada, for example, in one of its publications or in a research paper.
- (6) Thereafter, the Principal Investigator may, subject to subsection 6(5), carry out a secondary analysis, but such analysis shall be based solely on the approved “Proposed Output” produced pursuant to this contract and be related to the analytical work undertaken to produce the “Proposed Output”.
- (7) The Principal Investigator agrees to work with Statistics Canada in trying to meet the requirements of peer and institutional review required for the publication or research paper. For the Research Data Centres, a timetable for conducting the peer and institutional review is available in the guidelines for producing the “Proposed Output”.
 - In the event the “Proposed Output” fails a peer or institutional review and Statistics Canada decides not to publish it, Statistics Canada will give the Principal Investigator written notice to this effect within thirty days of having made the final decision.
 - Subject to subsections 6(5) and 10(2), in the event Statistics Canada notifies the Principal Investigator in writing that the proposed output will not be published, the Principal Investigator will not be prevented from:
 - (a) Publishing the “Proposed Output” elsewhere, and/or
 - (b) Using the “Proposed Output” for purposes of the attainment of an educational degree.

OWNERSHIP

6. (1) The microdata file and related documentation shall at all times be and remain the sole and exclusive property of Statistics Canada, it being mutually agreed that this contract pertains to the use of the microdata file and related documentation to produce a “Proposed Output” for Statistics

Canada and that nothing contained herein shall be deemed to convey any title or ownership interest in the microdata file or the related documentation to the Principal Investigator or the Co-investigators. The computer equipment provided for use by the Principal Investigator and the Co-investigators must never be removed from the premises of Statistics Canada and remains the sole and exclusive property of the access facility.

- (2) Statistics Canada reserves the right to publish in whole or in part, to publish an amended version or not to publish at all, as Statistics Canada deems appropriate, the “Proposed Output” produced by the Principal Investigator pursuant to this contract.
- (3) Copyright in the “Proposed Output” produced by the Principal Investigator pursuant to this contract shall vest in Her Majesty the Queen in Right of Canada. The Principal Investigator shall provide to Statistics Canada at the completion of the contract or at such other time as Statistics Canada may require a written permanent waiver of Moral rights from every author who contributed to the aforementioned material. Statistics Canada (Her Majesty the Queen in Right of Canada) hereby grants to the Principal Investigator a non-exclusive license to use, reproduce, publish and distribute the "Proposed Output" for any purpose, including, without limitation, research, teaching and publication in any medium.
- (4) Secondary releases of the “Proposed Output” may be considered by Statistics Canada subject to obtaining consent from the Principal Investigator.
- (5) In publishing the “Proposed Output” elsewhere, using the “Proposed Output” in the attainment of an educational degree or carrying out any secondary analysis, any reports, documents, or materials which are subsequently prepared by the Principal Investigator which use, incorporate or are in any way based on any material produced under this agreement, shall prominently display the following notice:

“The research and analysis are based on data from Statistics Canada and the opinions expressed do not represent the views of Statistics Canada.”

CONFLICT OF INTEREST

7. (1) All persons engaged in the course of carrying out this contract shall conduct themselves in accordance with the principles and spirit of the *Conflict of Interest and Post-Employment Code for the Public Service*.

(2) Should a conflict exist prior to the commencement of this contract or be acquired or develop during the life of this contract, the person with the conflict engaged in carrying out this contract shall be responsible for discussing the conflict with the Director of the Division sponsoring the project or the Manager of the Research Data Centre Project and, should it be determined that a conflict exists, for completing the Confidential Report as required by the Conflict of Interest and Post-Employment Code for the Public Service.

(3) No person engaged in the course of carrying out this contract may use any of the information gained by accessing the confidential data for any other purpose except that which was agreed upon in this contract.

(4) Notwithstanding subsection 7(3), it is understood that the principles of the *Conflict of Interest and Post-Employment Code for the Public Service* were not meant to prohibit the persons engaged in this contract from accomplishing any secondary analysis as permitted by the contract.

SECURITY REQUIREMENTS

8. (1) Any material to be removed from the Statistics Canada premises by the Principal Investigator or Co-investigators must first be screened by Statistics Canada to ensure that there is no risk of disclosure of confidential information or information that may lead to the identification of an individual respondent. It is the responsibility of the Principal Investigator or Co-investigators to take all precautions to avoid disclosure of confidential information or information that may lead to the identification of an individual respondent. The Principal Investigator or Co-investigators may remove summary files, tabulations and analytical output under the terms of this subsection.
- (2) The Principal Investigator and the Co-investigators shall not remove any of the original data set or copies of subsets of the microdata file or any confidential sensitive statistical information provided pursuant to this contract from the premises of Statistics Canada.
- (3) The Principal Investigator and the Co-investigators shall be provided with copies of all relevant Statistics Canada policies related to confidentiality, privacy and security and the standard operating procedures of the appropriate access facility and shall acknowledge in writing their compliance with all of the policies and operating procedures.

TERM

9. This contract comes into force when signed by both parties and shall continue in force until _____ unless revoked or terminated at an earlier date.

TERMINATION

10. (1) Where the Principal Investigator is in default in carrying out any of its obligations under this Contract, Statistics Canada may, upon giving written notice to the Principal Investigator, terminate the Contract immediately.
- (2) The Contract may, by providing 30 days written notice, be terminated by mutual written consent between the Principal Investigator and Statistics Canada.
- (3) Any notice to be given to Statistics Canada or the Principal Investigator shall be sent by registered mail to:
- (Address of Statistics Canada)* *(Address of the Principal Investigator)*
- (4) Any notice shall be deemed to be effective on the day it is received at the address set out above.

PENALTIES

11. (1) As a Deemed Employee of Statistics Canada, the Principal Investigator and the Co-investigators are subject to all the applicable penalties provided for in the *Statistics Act* for contravention of any of the confidentiality provisions and is liable on summary conviction to any of the applicable fines or imprisonment terms.
- (2) Subsection 11(1) survives indefinitely the completion of this contract or the termination of this Agreement pursuant to subsections 10(1) or 10(2).

AMENDMENT

12. No amendment to this contract shall be valid unless it is reduced to writing and signed by the Parties hereto.

CONSIDERATION

- 13. The Parties agree that consideration for this agreement shall be the mutual promises and covenants of the Parties included in this contract.

ENTIRE AGREEMENT

- 14. This contract constitutes the entire agreement between the Parties listed below and Statistics Canada with respect to the subject matter described herein and supersedes all previous negotiations, communications and other agreements on the same issue with Statistics Canada unless they are specifically incorporated by reference in this contract.

IN WITNESS WHEREOF, this Agreement has been executed in duplicate on behalf of Statistics Canada and the Principal Investigator by:

FOR STATISTICS CANADA:

| | | |
|---------|-------|--|
| _____ | _____ | _____ |
| Witness | | Date |
| | | Director of <i>(Name of Division)</i> or |
| | | Manager of the Research Data Centre |
| | | Program |

FOR THE PRINCIPAL INVESTIGATOR:

| | | |
|---------|-------|---|
| _____ | _____ | _____ |
| Witness | | Date |
| | | <i>(Name of Principal Investigator)</i> |

| | | |
|---------|-------|----------------------------------|
| _____ | _____ | _____ |
| Witness | | Date |
| | | <i>(Name of Co-investigator)</i> |

Appendix “A”

Research Proposal Title

Submitted By: Name of Division and Director
Name of Principal Investigator and Co-investigator
Address of Principal Investigator
Telephone Numbers
Affiliation

Proposal

Detailed description of the proposed research.

Data Requirements

List of files, additional data fields or variables required for the analysis.

Proposed Output

Indicate the type of output to be generated from the list below (delete those which do not apply)

- RDC working paper: A paper for the RDC working paper series authored by the Principal Investigator for Statistics Canada.
- Peer-reviewed journal article: A journal article authored by the Principal Investigator for a peer-reviewed journal.
- Book or Book chapter: A book or book chapter authored by the Principal Investigator.
- Thesis or Dissertation: A graduate level thesis or dissertation.
- Commissioned Report: A commissioned report authored by the Principal Investigator for _____ (insert name of commissioner and proposed title of the report).

Completion Date

Proposed Completion Date: (Enter Dates and expected product)

Research Location

Indicate where the Co-investigator is to be located for purposes of this project.

Source of Funding

List any agencies which may be providing funds related to this project.

APPENDIX 4 - THE OATH



Statistics Canada Statistique Canada

Note: This to be taken by all team members listed on the contract.

OATH OR AFFIRMATION OF OFFICE AND SECRECY

I, _____, do solemnly swear (or affirm) that I will faithfully and honestly fulfill my duties as an employee of Statistics Canada in conformity with the requirements of the Statistics Act, and of all the rules and instructions thereunder and that I will not without due authority in that behalf disclose or make known any matter or thing that comes to my knowledge by reason of my employment.

Sworn (affirmed) before me

at

this _____ day of _____ 20____.

(Authorised to administer oaths and affirmations pursuant to the Statistics Act)

Source: Statistics Act.

APPENDIX 5 – SAMPLE RESEARCH CONTRACT AMENDMENT

RESEARCH DATA CENTRES MICRODATA RESEARCH CONTRACT

AMENDMENT

(Could be used as an extension of contract, addition of team members or addition of dataset)

BETWEEN:

HER MAJESTY THE QUEEN IN RIGHT OF CANADA, represented by the Minister of Industry, designated as the Minister for purposes of the Statistics Act, (hereinafter referred to as "Statistics Canada"),

AND: (NAME OF INVESTIGATOR)
(hereinafter referred to as the "Principal Investigator")

(NAME OF CO-INVESTIGATOR OR RESEARCHER)
(hereinafter referred to as the "Co-investigator")

WHEREAS Statistics Canada and the Principal Investigator named above have signed an On-Premises Microdata Research Contract (copy attached) to undertake a study of -----;

AND WHEREAS the Co-Investigator named above will be working on this project as a 'deemed employee' of Statistics Canada and by signing this amendment will be subject to the same terms, obligations and penalties as outlined in the attached research contract;

OR

AND WHEREAS the Investigator wishes to extend the completion date of the contract from ----- to ----- in order to properly complete the research and analysis as outlined in the attached research contract;

OR

AND WHEREAS the Investigator wishes to have access to the following datasets:
- (Name of dataset or datasets)
as a new data set to be used for the data output in the original contract in order to properly complete the research and analysis as outlined in the attached research contract;

Project ID number

IN WITNESS WHEREOF, this amendment has been signed on the dates indicated below.

FOR STATISTICS CANADA

Witness Date Manager, Research Data Centres Program

FOR THE INVESTIGATOR

Witness Date Investigator (could indicate: signing for the team)

Witness Date Co-Investigator

Project ID number

APPENDIX 6 - SAMPLE DISCLOSURE REQUEST FORM

Research Data Centre Disclosure Request Form

| | |
|--------------------------|---------------|
| Name: | Date: |
| Email address: | Phone number: |
| Project number or title: | |

These few simple checks can facilitate the release of disclosure requests:

| Check that the following is complete: | Yes / Not Applicable |
|---|----------------------|
| Variables are clearly labelled, particularly recoded variables. * | |
| Cell sizes below the minimum requirement are collapsed or re-categorized. | |
| Mistakes in the output are corrected. | |
| Specific disclosure issues for the survey (e.g., rounding of EDS output) are addressed. | |

* If Applicable, please provide **variables information file**: S:\Project number\Disclosure\Pending\Folder name

Directory of files to be vetted: S:\Project number\Disclosure\Pending\Folder name

Output for Release:

| File Name | Survey Name and cycle(s) used | Weighting (e.g., population, fractional/adjusted/sample, bootstrap) | Types of Methods (Insert number(s) from list below or specify) | Subsample Characteristics (e.g., gender, age range, regional distribution, etc.) |
|-----------|-------------------------------|---|--|--|
| | | | | |

Project ID number

| File Name | Survey Name and cycle(s) used | Weighting (e.g., population, fractional/adjusted/sample, bootstrap) | Types of Methods (Insert number(s) from list below or specify) | Subsample Characteristics (e.g., gender, age range, regional distribution, etc.) |
|------------------|--------------------------------------|--|---|---|
| | | | | |
| | | | | |
| | | | | |

Types of Methods:

1. **Descriptive methods** (e.g., Frequencies, Cross-tabular analysis, means and distributions)
2. **Scaling methods** (e.g., Factor Analysis)
3. **Graphs** (e.g., histograms)
4. **Regression methods** (e.g., OLS, ANOVA, Logistic, Probit, Tobit)
5. **Complex methods of modeling** (e.g., Structural equation modeling, Hierarchical Linear Modeling, Growth analysis, Survival analysis, Event History Analysis, Simultaneous-equations Models, Fixed Effects Models, Random Effects Models)

Supporting Documentation Files:

Includes information to address disclosure issues but is not for release:

- **A clearly labelled un-weighted version of the tabular data.**
- **The necessary supporting cross-tabulations of binary variables for variance-covariance matrices.**
- **Clear indication of the source of the data if contextual information has been merged with the dataset.**

Project ID number

| Directory | File Name |
|--|-----------|
| S:\Project number\Disclosure\Pending\Folder name | |
| S:\Project number\Disclosure\Pending\Folder name | |

Residual Disclosure:

A re-release of the same output after slight modifications greatly increases the risk of residual disclosure. Statistics Canada strongly recommends that researchers submit as few versions of the output as possible for release from the RDC.

| Check: | Yes / No / Not Applicable |
|--|---------------------------|
| Has a version of this output has been previously released? | |
| Have variables been recoded or modified affecting only a few respondents since a previous release? | |
| Have individual outliers or cases have been dropped from the analysis since a previous release? | |
| Has the sub-sample or target population has changed slightly from a previous release? | |

If the answered is **'yes'** to any one of these questions, the output is at risk for residual disclosure. Please **consult with the RDC Analyst** for solutions and strategies.

Please list any other comments or concerns you may have for your disclosure request: