Preface

This guide applies to researchers working under contract to the federal department Statistics Canada accessing Statistics Canada data in the university Research Data Centres (RDCs). This guide is designed to identify researchers’ responsibilities, as well as the responsibilities of the RDCs staffs and Statistics Canada so that researchers understand all facets of the RDC. It is organized as follows:

♦ Chapter 1: Introduction to the Research Data Centres and the general application process in accessing the centres.

♦ Chapter 2: General Policies, privileges and responsibilities.

♦ Chapter 3: Procedures to follow at each phase of the research process, starting from the project proposal to submission of final papers.

♦ Chapter 4: Describes policy and procedures for protecting the confidentiality of respondents in statistical output and how to obtain release of statistical outputs.

♦ Appendices on data available in the RDCs, guidelines for disclosure analysis helpful checklists to follow for each phase of involvement with the RDC, and a note on the use of standard statistical software with surveys of complex design are included.

Any comments or questions can be directed to the RDC staff or contact Gustave Goldmann, Project Coordinator, Statistics Canada (by phone: 613-951-1472 or by e-mail: gustave.goldmann@statcan.ca).
# Table of Contents

**PREFACE**

**TABLE OF CONTENTS**

**CHAPTER 1. INTRODUCTION TO THE RESEARCH DATA CENTRES AND APPLICATION PROCESS IN ACCESSING THE CENTRES**

1.1 Scope .................................................................................................................................................. 5
1.2 Application process ................................................................................................................................ 5
1.2.1a Research proposal .......................................................................................................................... 5
1.2.1b Proposal review process .................................................................................................................. 6
1.2.2 Security clearance ............................................................................................................................. 7
1.2.3 Contract .......................................................................................................................................... 7
1.2.4 Orientation ....................................................................................................................................... 7
1.2.5 Oath .................................................................................................................................................. 7

**CHAPTER 2: POLICIES, PRIVILEGES AND LEGAL RESPONSIBILITIES**

2.1 Confidentiality of Statistics Canada Data ............................................................................................. 8
2.2 The Legal Requirement to Maintain Confidentiality of Data at an RDC ................................................... 8

**CHAPTER 3: THE LIFE CYCLE OF RESEARCH AT THE RDC**

3.1 Beginning your project at the RDC ......................................................................................................... 9
3.1.1 Physical security of the centre ............................................................................................................ 9
3.1.2 Computer facilities and security ....................................................................................................... 9
3.2 During your stay at the RDC .................................................................................................................. 9
3.2.1 Maintaining facilities security .......................................................................................................... 9
3.2.2 Maintaining workstation security .................................................................................................... 10
3.2.3 Maintaining data security .................................................................................................................. 10
3.2.4 Maintaining confidentiality and obtaining release of research output ............................................. 10
3.2.5 Documenting your research ............................................................................................................. 12
3.2.6 Progress reports .............................................................................................................................. 12
3.2.7 Getting Support ............................................................................................................................. 12
3.2.8 Software packages for complex-designs survey data .................................................................. 12
3.2.9 Obtaining new data or carrying out a different research project .................................................. 13
3.2.10 Add or remove a team member .................................................................................................... 13
3.3 Completing the contract ....................................................................................................................... 14
3.3.1 Producing an RDC working paper ................................................................................................ 14
3.3.1a Peer review ................................................................................................................................... 14
3.3.1b Institutional review ....................................................................................................................... 14
3.3.2 Paper submission ............................................................................................................................. 14
3.3.3 Producing subsequent discussion papers and journal articles ...................................................... 14
3.3.4 Seminars and conference presentations .......................................................................................... 15
3.4 Exit process .......................................................................................................................................... 15
3.4.1 Provide output documentation ....................................................................................................... 15
3.4.2 Data and programs archiving ......................................................................................................... 15
3.4.3 Returning badge or access keys ..................................................................................................... 15
3.5 After exit ............................................................................................................................................... 15
3.5.1 Returning for follow-up work ......................................................................................................... 15
3.5.2 Re-entering the RDC on a new agreement ..................................................................................... 16
3.5.3 Related papers and research reports ............................................................................................... 16
CHAPTER 4. DATA ACCESS, CONFIDENTIALITY, DISCLOSURE ANALYSIS POLICIES AND PROCEDURES.................................................................17

4.1 WHAT IS DISCLOSURE? .............................................................................................................................................................................17
  4.1.1 Types of Disclosure ........................................................................................................................................................................17
  4.1.2 Some examples of disclosure in survey data .................................................................................................................................17
  4.1.3 How do we protect against disclosure? ...............................................................................................................................................18

4.2 HOW TO TAKE RESEARCH OUTPUT FROM THE RDC .....................................................................................................................18
  4.2.1 Tracking Requests to Clear Research Output ...................................................................................................................................19

4.3 DISCLOSURE ANALYSIS FOR TABULAR OUTPUT .........................................................................................................................20

4.4 DISCLOSURE ANALYSIS FOR PARAMETRIC MODEL OUTPUT .........................................................................................................20

4.5 DISCLOSURE FOR NON-PARAMETRIC, SEMI-PARAMETRIC OR HIERARCHICAL DATA .................................................................21

4.6 DISCLOSURE ANALYSIS FOR VARIANCE-COVARIANCE MATRICES .............................................................................................21

4.7 SOME QUICK TIPS FOR HOW RESEARCHERS CAN SPEED UP THE PROCESS; .............................................................................21

4.8 QUALITY GUIDELINES FOR RELEASING DATA ...............................................................................................................................21

4.9 DISAGREEMENTS.........................................................................................................................................................................................22

APPENDIX 1. DESCRIPTIONS OF DATABASES HOUSED IN THE RDC .................................................................................................23

  SURVEY OF LABOUR AND INCOME DYNAMICS (SLID) ..............................................................................................................................23
  THE NATIONAL LONGITUDINAL SURVEY OF CHILDREN AND YOUTH (NLSCY) ..............................................................................................23
  NATIONAL POPULATION HEALTH SURVEY (NPHS) .................................................................................................................................24
  THE WORKPLACE AND EMPLOYEE SURVEY (WES) .................................................................................................................................25
  YOUTH IN TRANSITION SURVEY (YITS) ....................................................................................................................................................26

APPENDIX 2. MORE ON DISCLOSURE AND DISCLOSURE RISK ...........................................................................................................27

APPENDIX 3. ON-PREMISES ACCESS MICRODATA RESEARCH CONTRACT .............................................................................................33

APPENDIX 4. PROPOSALS GUIDELINE AND LAYOUT .............................................................................................................................39

  APPENDIX 4-A. SAMPLE LAYOUT OF A PROPOSAL ..............................................................................................................................40

APPENDIX 5. THE OATH ..................................................................................................................................................................................41

APPENDIX 6. RESEARCH CONTRACT AMENDMENT ............................................................................................................................42

APPENDIX 7. SAMPLE OUTLINE OF A RESEARCH PAPER .....................................................................................................................43
Chapter 1. Introduction to the Research Data Centres and application process in accessing the centres

Research Data Centre (RDC) is a combination of a secure Statistics Canada office on a university campus and a statistical analysis computer laboratory. As a Statistics Canada office, the RDCs hold copies of confidential social microdata. The creation of RDCs in major universities across Canada offers a unique opportunity to extend the analyses of these data sets beyond the physical boundaries of Statistics Canada offices while maintaining strict standards of confidentiality.

1.1 Scope

The proposal for the RDCs arose from a report of a joint-committee of Social Sciences and Humanities Research Council (SSHRC) and Statistics Canada, known as the Canadian Initiative for Social Statistics. To promote quantitative research and policy analysis in Canada, the idea of RDCs was recommended. The RDCs are to provide researchers with access to confidential longitudinal data collected by Statistics Canada. Confidentiality is to be ensured by housing copies of data files in physically secure location in various university campus across Canada managed by a Statistics Canada full-time employee. The research findings are to be disseminated to the public and the policy community.

The success of this endeavour is in the best interests of all concerned; Statistics Canada, Social Sciences and Humanities Research Council, the institutions in which the centres are located, the social science research community, other federal government agencies and Canadian society.

Access to the confidential data in the RDCs is open to any researcher (or team) from public organizations with legitimate research interests related to public policy development or fundamental research (e.g. universities, non-government organizations, provincial governments, federal government agencies and think tanks), upon successful completion of the application process.

1.2 Application process

The confidential data kept by Statistics Canada in a RDC are protected under the Statistics Act of Canada. Only employees or deemed employees of Statistics Canada may have access to those data. A process has been established to permit researchers with approved proposals to work with these data. Researchers must become deemed employees of Statistics Canada and produce a research paper that follow the Statistics Canada’s mandate (Statistics Act section 3, 19, 20, 22). Thus, to granted access to the RDCs, a researcher is to submit a research project proposal, undergo a security evaluation, and take an oath to protect confidentiality. The proposed project will become the contracted project between the researcher and Statistics Canada.

1.2.1a Research proposal
Research proposals are submitted to an adjudicating committee operating under the auspices of the SSHRC and Statistics Canada. It is done electronically by accessing the submission form on the SSHRC Web site:

http://www.sshrc.ca/rdc/english/applicants.html

Researchers must submit a separate proposal for each new project they wish to carry out at 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. See Appendix 4-A for proposal layout.

1.2.1b Proposal review process

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

The committee will evaluate each proposal based on its scientific merit and recommend to Statistics Canada whether or not the project meets the criteria that were established. As such, the task of the committee differs fundamentally from other selection committee run by SSHRC, that it is not overseeing a competition. The committee does not select winners from a range of applicants, and there is no financial grant involved in this process.

Criteria for evaluation:
• Scientific merit of the proposed research (the Statistics Canada member of the peer review committee will also assess whether the project falls within the agency’s mandate).

• Viability of the methods to be applied, given the data on which the analysis will be performed.

• Evidence of the ability of the principal investigator(s) to carry out the project (curriculum vitae).

• Demonstrated need for access confidential microdata.

• Confidentiality review - proposal will be reviewed by Statistics Canada to ensure that issues of confidentiality and disclosure avoidance have been addressed. Direct interaction with the researchers may occur at this stage. For instance, if it appears that the output from the project would result in disclosure, staff from Statistics Canada will work with the researchers to address the issue.

The review process will produce one of two outcomes – acceptance or rejection; and the decision must be unanimous among all peers. The applicant will be advised of the results. The decision rationale will be included in the report of the evaluation results for the projects that are rejected.

The evaluations of the projects will be conducted on a schedule that generally ensures that decisions are returned to the applicants within two months of submission. However, instances in which the results of the review are not unanimous may take longer.
1.2.2 Security clearance

After the project proposal has been accepted and approved, all members of the proposed research team will agree to undergo an enhance reliability security check. The necessary forms and instruction can be obtained through your local RDC analysts. (Note: foreign or non-Canadian researchers should provide a photocopy of their passport, student visa/work permit, and a brief written explanation of their role in the research project.)

1.2.3 Contract

The accepted proposal will become the contracted project between the researcher and Statistics Canada. The agreement will specify the following items:

• Project plan (a copy of the proposal).

• Date sets to be provided by Statistics Canada.

• Computer hardware and software, and other related services to be provided by the hosting university, and, if appropriate, provided by the researcher.

• Project schedule detailing the period and intensity of use of the RDC facilities.

• Agreement of the researchers to abide by the RDC security and confidentiality requirements.

• Agreement of the researchers to provide progress reports and other administrative reports.

Please note the contract will give researchers access only to the microdata specified in the approved research project and allows access to data only for the purpose of completing that proposal. See Appendix 3 for a sample contract.

1.2.4 Orientation

The last element of the application progress is the orientation session. It is a presentation designed to familiarize researchers with the privilege and duty of becoming deemed employees of Statistics Canada. It includes discussion on policies on data confidentiality, disclosure analysis, and the other information regarding the operation of RDC. Procedures for obtaining release of research output will also be discussed.

1.2.5 Oath

Before accessing data, researchers you will be asked to take an Oath of Office (oath of secrecy) to become the deemed employee of Statistics Canada. This is the legal requirement to maintain data confidentiality. The oath is binding for life beyond the end of the research contract; it will be re-affirmed every two years if required. A copy of the oath can be found in Appendix 5.
Chapter 2: Policies, Privileges and Legal Responsibilities

Polices, privileges and legal responsibilities that apply to researchers accessing confidential Statistics Canada data in a Research Data Centre will be outlined in this chapter.

2.1 Confidentiality of Statistics Canada Data

Statistics Canada takes great care to respect the trust that all respondents place in the organization to safeguard the privacy and confidentiality of the information that they provide to the agency. It is this trust of the Canadian that makes it possible for Statistics Canada continues to collect accurate and meaningful information. Almost all of the household surveys carried out by Statistics Canada do not have mandatory participation; that is respondents volunteer their time and information freely, and are not compensated with money. In return, Statistics Canada goes to great lengths to protect the confidentiality of our respondents.

2.2 The Legal Requirement to Maintain Confidentiality of Data at an RDC

In order to carry out research in the RDC, researchers will be granted access to confidential statistical data (also called sensitive statistical information or information provided in confidence). Confidential statistical data are safeguarded by the application of various levels of both physical and electronic security measures.

Before accessing data, researchers will be asked to take an Oath of Office to become a deemed employee of Statistics Canada. As part of this process they agree 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 means that researchers are personally accountable to uphold the confidentiality provisions of the Statistics Act, and not to reveal anything about individual identifiable 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. Much of what is discussed in this document is intended to help prevent an accidental breach of confidentiality.

It also is 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 research centre indeterminately. Statistics Canada employees are subject to the same consequences and penalties.
Chapter 3: The Life Cycle of Research at the RDC

3.1 Beginning your project at the RDC

3.1.1 Physical security of the centre

The RDC office is protected by a security system (e.g. electronic card key, combination lock). After your RDC application process has been completed, you will be given the proper procedures for entry to the centre during the orientation. It will enable you to access the RDC only during operating hours. To help us maintaining the physical security of the place, researchers are asked not to share their entrance information or bring visitors into the centre.

3.1.2 Computer facilities and security

The computer setup will typically include workstations running Windows NT. The server environment is not connected to an outside network such as the Internet or the university network. These workstations will not have write access to diskettes or CDs, nor it equipped with printing capacity. All files transfers and printing will be done through the RDC analyst.

Researchers are not assigned specific workstations in the RDC laboratory; researchers are asked to approximate schedule for access to the computer in consultation with the RDC analyst. It is important to book a workstation in advance to minimize conflicts over resources and workstations. As a courtesy to other users, please adhere to the booking as much as possible.

3.2 During your stay at the RDC

3.2.1 Maintaining facilities security

- Ensure that the security clearance and entrance information such as identification, key code, card-lock combination, door key etc. are current and appropriate. If you suspect there is a breach in security or lost your key, inform the RDC analyst immediately.

- RDC entrance information must not be disseminated.

- Ensure RDC office door is closed at all times.

- Please sign in every time when working in the RDC. Sign-in procedures may vary across RDC locations but a record of attendance will be taken.

- Visitors are not permitted into the secure RDC facilities.

- Adhere to any further requirements specific to the physical site location.
3.2.2 Maintaining workstation security

- 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, beepers or similar device may be used inside the RDC.

- Researchers are accountable for computer use, through passwords and system logs. Do not share passwords.

- Log off after every work session. If you need to be away from your workstation during the work session, please save all your work and lock the computer.

- Researchers are asked to save all work in an assigned folder for backup and system maintenance.

3.2.3 Maintaining data security

- Do not discuss confidential information at any time outside the RDC, whether over the telephone or in person.

- Please place any output containing confidential data that the researchers wish to discard into the designated receptacle or return to the RDC Analyst.

- Do not ask other researchers at the RDC for access to their data sets. Researchers are granted access only to the data approved for use in the approved research project. Written requests or a new proposal is required for access to other data sets (see section 3.2.9).

- Do not give other researchers access to any data sets without permission from the RDC analyst. If the principal investigator wishes to give other RDC team members access to secondary datasets\(^1\), it needed to be granted by the RDC analyst.

- Do not carry out analyses for colleagues or other researchers who do not have approved RDC research projects.

- Do not use RDC data files to conduct data analysis outside the mandate of the approved project.

3.2.4 Maintaining confidentiality and obtaining release of research output

Users should be aware that Statistics Canada considers important not only to avoid disclosure of confidential information, but also to avoid the perception of disclosure. It is the trust of our respondents that makes data collection possible.

---

\(^1\) A secondary dataset is one derived from the micro data files for analytical use.
The following is a brief summary of our policies regarding confidentiality and obtaining release of the research output.

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

- Researchers must meet with the RDC analyst at the beginning of their project to discuss data confidentiality, security policies and security procedures. This meeting can be set up as part of the RDC orientation session.

- All output leaving the RDC must undergo disclosure analysis. Disclosure analysis is the means by which researchers and the RDC analyst ensure that any material removed from the RDC does not pose a disclosure risk of data confidentiality. Please ensure there is sufficient time for the analysis to be completed. See Chapter 4 and Appendix B for a complete discussion of the topic.

- Researchers are to provide documentation of data sets and programs used in producing output to help the RDC analyst better understand the materials. Working interactively with the person performing the disclosure analysis will help to avoid misunderstandings and speed up the approval process.

- 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 other project researchers with deemed employee status.

- We encourage research projects that emphasize model output rather than tabular (descriptive) output to reduce disclosure risk.

- We discourage removal of “intermediary output” from the RDC. Intermediary output is typically produced during the exploratory and model definition stages of analysis. They tend to consist of such things as detailed tables of descriptive statistics. Such outputs often present more detailed information than final model outputs; their removal can sharply increase the risk of disclosure, especially residual disclosure.

Within the secure premises of the RDC, output of any type can be printed onto paper, except it must not be removed from the RDC. Output without disclosure control will be printed onto coloured paper to avoid accidental mix up with outside material brought into the centre. Researchers can access those materials during any working sessions in the RDC. At the end of the session, materials should be returned to RDC analysts for locked up. Those materials will be shredded upon researcher’s request or when the project is completed.

Exploration of data is highly encouraged, however, researchers must comply with Statistics Canada rules and regulations on data confidentiality when removing output from the RDC.
3.2.5 Documenting your research

Researchers are asked to document work often and in an orderly manner. At the end of each research project, researchers are asked to provide secondarily created datasets, model programs, relevant documentation, and output used in the writing of the research paper.

3.2.6 Progress reports

Periodically the RDC analyst will ask each research team to fill in a brief progress report on the project. It is to help us on resource and usage planning with the RDC facilities.

3.2.7 Getting Support

Researchers can direct questions about the dataset to the RDC analyst. The analyst will either have the answer for question or else direct it to other Statistics Canada employees. This type of knowledge exchange benefits both the researchers and the Statistics Canada data programs when managed properly.

3.2.8 Software packages for complex-designs survey data

Most 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 does exist that will account for the design complexities.

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.

While there are different accepted approaches to accounting for the design complexities when analyzing survey data, the approach being recommended here is a design-based approach. In this approach, survey weights are used to produce estimates of parameters. Other design details, such as information about stratification, clustering, and weight adjustments, are required for the production of design-based variance estimates needed for inference.

Several software packages have been developed for implementing a design-based approach for many different analytical procedures. The packages vary in what analytical procedures they contain and in what methods they use for design-based variance estimation. The package that best matches
the needs of the analysis and the design information available from a particular survey would have
to be chosen.

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.

There has been considerable research carried out and continues to be carried on the appropriateness
of various statistical analysis methods using complex-designs survey datasets. The Data Analysis
Resource Centre (DARC) can provide suggestions on suitable methods and software tools. DARC
also conducts seminars and research on various methodological problems particularly of the
longitudinal surveys. These services are offered on a cost-recovery basis. For more details, please
contact Georgia Roberts in Ottawa by phone at (613) 951-1471 or by e-mail at
georgia.roberts@statcan.ca.

3.2.9 Obtaining new data or carrying out a different research project

Researchers may find that while working on the approved projects, there maybe a situation when
they wish to use a different data set as proposed in the original proposal, such as an earlier cycle of
the survey or a different survey entirely. For situations as mention, the following policies will
apply:

• To obtain access to different data, researchers should submit a data request form (which can be
  obtained from the RDC Analyst) and a memorandum to the RDC analyst that describes the
  required data set. Include in the memorandum reasons for the request; how the new data would
  fit the original research project, how its use may affect the schedule, and how it may affect the
  research output to be removed from the RDC.

• If the proposed use of the new data does not fit within the original proposal, researchers asked to
  begin a new project. This proposal may be subject to the formal project review process.

• Do not start a new project without first informing the RDC analyst. Violation of this rule will
  result in immediate withdrawal of data access privileges including data access to projects
  currently under contract.

3.2.10 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. To add a new researcher, please provide a written memorandum along with a
curriculum vita of the researcher to the RDC analyst. The new member will be required to undergo
the enhance reliability security check, take the oath of office and attend the orientation session.
3.3 Completing the contract

Statistics Canada has a mandate to provide a publicly available document from all research conducted on its data sets. Researchers are responsible for producing this report from their research. See contract, section (5), “Limitations on uses of the microdata file and proposed output”.

3.3.1 Producing an RDC working paper

The contract (the memorandum of understanding) in most cases, the author(s) of the research project has agreed to produce a paper describing the proposed researcher project, and provide the research findings and outputs to Statistics Canada. This paper must meet the requirements of both peer and institutional reviews. Researchers should restrict themselves to interpretive statements supported by the statistical analysis and remain policy neutral. See Appendix 8 for guidelines in writing this report.

3.3.1a Peer review
The peer review process is designed to evaluate the paper from a scientific and methodological perspective. Each paper will go to two reviewers with specifications and guidelines for the peer review. The reviewers will remain unknown to the author(s).

3.3.1b Institutional review
The institutional review is conducted to ensure that the paper does not contain policy comments or other statement that are outside of Statistics Canada's regulations and policies. This step will be done simultaneously with the Peer review process.

Researchers are responsible for addressing the comments of the reviewers and submitting a final draft of the paper to Statistics Canada and will have two months to do so.

3.3.2 Paper submission

Papers ready for peer review and final revised drafts can be submitted through the RDC to:
   Gustave Goldmann
   Room 1710 Main Building
   Tunney’s Pasture
   Statistics Canada
   Ottawa, ON K1A 0T6
   Phone: (613) 951-1472

3.3.3 Producing subsequent discussion papers and journal articles

The researcher is free to produce subsequent papers based on the same statistical analysis. These papers may contains policy commentary or other interpretive statements, and be published anywhere author sees fit. However, the confidentiality requirements that applied to the first paper remain in effect for all subsequent publications.
3.3.4 Seminars and conference presentations

If they wish, researchers may present their findings at seminars and conferences. All work presented from your project must be based on analysis reported in the paper for Statistics Canada, so that it may undergo disclosure analysis and be revised according to the peer review process.

3.4 Exit process

3.4.1 Provide output documentation

When producing the final paper, researchers are asked to provide secondarily created data sets, model programs, relevant documentation and output used in creating the research report or paper. This information may be part of the research output clearing process for release.

You are asked to provide:

- The variables, the model program, the log (if the statistical software produces a log) to accompany each tabulation or model output.

- A brief description (1 or 2 sentences) of both the following:
  - The data contained in a secondarily created data set; and
  - Variables that have been derived for your analysis, including the value labels and cell counts where applicable.

3.4.2 Data and programs archiving

The research account will remain active for at least six months following the completion of a Statistics Canada contract before being archived.

3.4.3 Returning badge or access keys

Researchers must return their identity badge or RDC laboratory access key upon project completion. Researchers are also asked to return their access badge or key if they plan to be absent from the RDC for periods longer than two months.

3.5 After exit

3.5.1 Returning for follow-up work

We allow researchers to return for reasonable amounts of follow-up work. The typical situation involves carrying out additional model estimations in response to journal referees’ comments.
The of the following is what we consider a *reasonable* amount of work:

- No new data or research samples will be used;
- Only a small amount of output is required that are very similar to output previously released.
- New requested output will not cause disclosure problems.
- The scope of the work is not significantly different from the original project, nor it is related to other research project.

### 3.5.2 Re-entering the RDC on a new agreement

Researchers will follow the usual application process by submitting a research proposal to the Social Science and Humanities Research Council (SSHRC) web site, as they did the first time they wanted access to data.

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.

### 3.5.3 Related papers and research reports

Researchers are invited to provide a copy of the abstract or the full document of any subsequent papers produced from the original RDC paper. Statistics Canada hopes to maintain an electronic library of publications resulting from the RDC program available to the social research community when appropriate.

Papers can be sent to the RDC analyst or to:

Dr. Gustave Goldmann  
Room 1710 Main Building  
Tunney’s Pasture  
Statistics Canada  
Ottawa, ON K1A 0T6  
E-mail: gustave.goldmann@statcan.ca
Chapter 4. Data access, confidentiality, disclosure analysis policies and procedures

The material in this chapter applies only to output that you intend to remove from the secure RDC site. Any amount of output can be produced and stored (up to available capacity) for use on-site. This chapter presents policies and procedures you must follow to clear your research results for release. Appendix B gives some 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 staff and researchers will work together to find solutions to data confidentiality problems.

4.1 What is disclosure?

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

4.1.1 Types of Disclosure

There are three types of disclosure.

Identity disclosure occurs when an individual can be identified from the released data, 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.

Care must be taken to examine all data 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. When released information can be combined to obtain confidential data this is called residual disclosure (e.g., suppressed data in one table can be derived from other tables).

4.1.2 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 were 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 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.

4.1.3 How do we protect against disclosure?

Disclosure analysis is the term applied to examination of output you 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. Appendix B presents some examples of specific problems and approaches to them.

The most important policy is this:

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

4.2 How to Take Research Output from the RDC

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

To track requests to clear research output; we must have information about the overall pattern of clearances (and requests for clearance) and about the supporting information that allows us to clear each request. To do this, please follow the procedures specified below.

1. *Project Clearance Directory.* You will be assigned a project computer directory where you should store all data sets, programming, logs, and output.

2. *Dated clearance request subdirectories.* These dated subdirectories of the project disclosure analysis directory will contain documentation of each specific release request for the project, the research output to be cleared, and disclosure output that allows us to clear the research output.

When you have research output you wish to remove from the RDC:

1. Make an appointment with the RDC analyst to discuss the clearance request.

2. Please provide:
   - The name of subdirectory under the project’s main clearance directory.
   - State any file naming conventions that will make it easier for the RDC analyst, HRDC, or Statistics Canada to revisit the clearance information after the project ends.
   - A description of the research sample – selection criteria and how the research sample differs from the original survey sample.
   - A glossary of the variable names and descriptions.
   - In properly documented output, *all variables should be LABELED and all variables should be in a consistent format* – e.g., identical variables should have the same names and the same formats across data sets. All tables and output should have clear titles indicating the nature of the output and whether the analysis uses weighted or un-weighted data (and the name of the weight variable). The programs that produced the material are part of the documentation.

3. Be prepared to discuss the clearance request with the person performing disclosure analysis. This may require you to make changes in your output to satisfy clearance requirements. Working interactively with the person performing disclosure analysis will minimize learning on both sides, avoid costly misunderstandings, and speed the review process. The clearer and more complete your request is, the more quickly it will be released.

The following sections describe how disclosure analysis is done for different types of output.
4.3 Disclosure Analysis for Tabular Output

Tabular output is what is often called ‘descriptive statistics.’ We ask that you limit the printing of your tabular output to the minimum necessary to describe the sample used in your models, and how it might compare to an underlying population.

1. 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 cells.
   - 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 values, and, sometimes, 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 we ask that you 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 quantiles of distributions should be discussed with your RDC analyst.

2. If you or the RDC analyst finds disclosure risk, we may ask you to modify the output (e.g., collapse categories). This will avoid disclosure risk at the expense of output detail, and is the preferred course of action.

4.4 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 recreation of simpler statistics such as means and totals for cross-classes of variables. This is particularly true of so-called “saturated models”, 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 you or the RDC analyst finds disclosure risk, you may be asked to modify your output and possibly your model specification. For example, you 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). We may ask you to report that the coefficient is in a certain range (e.g., positive and
significant). We will work with you to ensure, as much as possible, that your research results remain meaningful after any such modifications.

4.5 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. So too may estimates that account for the hierarchically organized or clustered nature of the complex survey data to be housed at the RDCs. An example is the now increasingly popular hierarchical linear model. This is likely if the clusters contain small numbers of observations. Special care must be taken with output such as these. Your RDC analyst will work with you to avoid disclosure risk in these cases.

4.6 Disclosure analysis for variance-covariance matrices

While we prefer that all data analysis be completed in the RDC, we realize some researchers will prefer to leave the RDC with variance-covariance matrices in order 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 2 for guidelines.

4.7 Some quick tips for how researchers can speed up the process:

- Correct any mistakes in your output and ensure that it is the final version of what you want to take out.
- Provide the un-weighted equivalents of tabular output.
- Ensure that all cells in your un-weighted tabular data contain at least five cases.
- Demonstrate that descriptive statistics are based on at least five cases and do not contain maximums and minimums for sensitive variables (especially income).
- Check with your RDC Analyst to ensure that you have provided the necessary supporting information for model output (i.e., cross-tabulations for saturated models).

See Appendix 2 for further instructions on approaching disclosure problems.

4.8 Quality guidelines for releasing data

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 might present problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used.

Each survey documentation contains guidelines on releasing acceptable estimates. Before releasing and/or publishing any estimate, users should first determine the quality levels of the estimate. The quality levels are acceptable, marginal and unacceptable. Data quality is affected by both sampling and non-sampling errors. The quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation. 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 you wish to analyze for more information on survey weights and data quality.

4.9 Disagreements

If you disagree with decisions, you are asked first bring your concern to the person who performed the disclosure analysis. If disagreement cannot be resolved, you can appeal to the RDC Project Coordinator at Statistics Canada, Gustave Goldmann at (613) 951-1472. Further appeals will be referred to the Statistics Canada directors associated with each survey.
Appendix 1. Descriptions of Databases Housed in the RDC

Survey of Labour and Income Dynamics (SLID)

SLID has added a new dimension - the changes experienced by individuals through time - to existing survey data on labour market activity and income. At the heart of the survey's objectives is the understanding of the economic well being of Canadians: what economic shifts do individuals and families live through? And how does it vary with changes in their paid work, family make-up, receipt of government transfers, or other factors? The survey's longitudinal dimension makes it possible to see such concurrent and often related events. SLID will be the first household survey ever to provide national data on the fluctuations in income that a typical family or individual experiences through time, which will give greater insight on the nature and extent of poverty in Canada.

SLID, which began in 1993, follows the same respondents for six years. A second “panel” was introduced in 1996 overlaps the first panel for a three-year period. Subsequent panels are introduced every three years.

Each panel includes about 15,000 households, including about 30,000 adults. A preliminary interview takes place at the beginning of each panel to collect background information. Each of the six years has a split-interview format, with labour topics covered in January and income topics in May. In both cases, questions refer to the previous calendar year. The income interview occurs in May to take advantage of income tax time when respondents are more familiar with their records. In addition, many respondents avoid the income interview by giving us the permission to consult their income tax file.

SLID will help answer questions such as these: what are the fluctuations in people's labour, income or family characteristics at the micro level? What events tend to coincide? How often do people change jobs or get laid off, with what impact on their total family income? How many families split or join together in a given time period? What proportion of households are persistently poor year after year, and what makes it possible for others to emerge from periods of low income?

With SLID, the focus extends from static measures to the whole range of transitions, durations, and repeat occurrences of people's financial and work situations. Since their family situation, education, and demographic background may play a role, the survey has extensive information on these topics as well. To ensure the sample remains representative, the window on each panel of respondents is kept to 6 years. A new panel of longitudinal respondents is selected every three years, so there is always an overlap between two panels of respondents.

The National Longitudinal Survey of Children and Youth (NLSCY)

Under the federal government's "Brighter Futures" initiative, Human Resources Development Canada (HRDC) and Statistics Canada have developed the NLSCY. The survey's purpose is to
generate information for policy analysis and program development on critical factors affecting the
development of children in Canada.

Beginning in the winter 1994-95, the NLSCY has been collecting data on approximately 25,000
children, ranging in age from newborn to 11 years. Participating households are selected from
Statistics Canada's Labour Force Survey sample frame. The NLSCY will be repeated at two-year
intervals to follow the children surveyed in 1994-95 as they grow to adulthood. The primary
respondent in the initial collection cycles is the household member most knowledgeable about the
child - usually the mother. She provides information about the parent(s) and children. In addition,
several interviewer-administered assessments to measure cognition are conducted in the home.
Children over the age of 9 fill out a questionnaire on their own, in the home. When children reach
the age of 16, they become the principal respondent for the survey. Provided the signed consent of
a parent is given, information is collected from teachers and principals by means of self-completed
questionnaires mailed to the school. Teachers also administer a short mathematical skills test in the
school to children in the sample who are in grades 2 to 10. These multiple sources of information
on the child will enrich the data and enhance data analysis.

The survey instruments for the NLSCY have been developed in consultation with an Expert
Advisory Group, subject matter specialists, federal and provincial/territorial officials. The survey
covers a broad range of characteristics and factors affecting child growth and development. It
collects information on the child's parent(s) and other family members as well as the characteristics
of the family and the neighbourhood. Additionally, the NLSCY gathers data on the child's health,
development, temperament, behaviour, child care and school experiences, relationships,
participation in activities, and family and custody history.

National Population Health Survey (NPHS)

The NPHS program started its first 12-month cycle of data collection in early June 1994 and is
continuing every second year. This initiative is the major component of a concerted effort to
improve the information available to support the development and evaluation of health policies and
programs in Canada during a time of economic and fiscal pressures on the health care system.
Extensive and invaluable consultations took place with representatives from Health Canada and the
provincial and territorial ministries of health in the development of the survey. As well, ad hoc
groups of researchers and specialists were convened to address content and methodological
problems. The survey will produce both periodic cross-sectional information to monitor programs
and longitudinal data to improve our understanding of the determinants of good health.

The NPHS targets three populations:

1) Household residents in all provinces (with the principal exclusion of populations on Indian
Reserves, Canadian Forces Bases and some remote areas in Quebec and Ontario);
2) Long-term residents expected to stay longer than six months in health care institutions with
four beds or more in all provinces; and
3) The Northern population including household residents in the Yukon and the Northwest Territories (with the principal exclusion of populations on Indian Reserves, Canadian Forces Bases and some of the most remote areas of the Territories).

The household and territorial components collect most of the information from a single household member. Interviewing one respondent simplifies the longitudinal follow-up. However, each time the respondent is re-surveyed, demographic, socio-economic and basic health-related information will also be collected from all members of the household in which he or she is then living. For the 1996-97 cycle, the household component of the NPHS started using the 17,276 longitudinal respondents from 1994-95. In addition, the provinces of Ontario, Alberta and Manitoba purchased supplemental samples so that a larger sample of dwellings could be selected. The final response rate was approximately 83% of households.

The Workplace and Employee Survey (WES)

The WES, a new undertaking by Statistics Canada, HRDC, and other federal government partners, is designed to provide information on how businesses manage technological and organizational change in the face of increasing international competition. The survey covers a broad range of issues relating to employers and their employees. It is unique in that employers and employees will be linked at the microdata level: employees are selected from within sampled business locations. Thus information from both the supply and demand side of the labour market will be available to enrich studies focused on either side of the market. WES data will provide valuable information to address the following policy priorities:

- skill development of employees;
- economic growth;
- family-friendly workplace practices;
- work arrangements and economic security;
- employment earnings and equity; and
- small business practices.

The WES has been under development for five years. The development process included a number of field tests and a large-scale pilot project. Development culminated with the first production survey, which went into the field early in the summer of 1999. Beginning with the initial 1999 sample, the survey will follow business locations for at least five years. The sample will be refreshed every two years to account for business closures in the original panel. Employees will be followed for two years, with the employee sample being completely redrawn whenever the location panel is refreshed. This methodology will result the following:

- the availability of representative estimates for locations and employees every second year;
- an ongoing panel of business locations for longitudinal research;
- rotating panels of employees that provide information on transitions from one employer to another and on the ability to include employer characteristics in labour analysis; and
- a new piece of statistical infrastructure to support special surveys and other information production activities.
The sample will include approximately 6,400 business and more than 20,000 employees, with microdata being available in late 2000.

**Youth in Transition Survey (YITS)**

The longitudinal Youth in Transition Survey makes it possible to track the development of the same people over many years. They will be observed in a number of different areas, such as the pursuit of further education, the integration into the labour force, and families and other relationships.

The survey will provide more precise data on commitment to schooling, the field of studies and drop out rates, workload, job preparation programs, activities, obstacles to education, stressful experiences, professional aspirations, family relations, living conditions and training, among others.

The survey will start officially in 2000 and will be a continuation of the two-year pilot project. It will consist of two cohorts: 20,000 18-to-20-year-olds surveyed in January; and 30,000 15-year-olds, surveyed in the spring.

In the spring of 2000, 32 nations will participate in the Organization for Economic Co-operation and Development’s (OECD) Program for International Student Assessment (PISA) to measure the reading, science and math skills of 15-year-olds and draw a link between their socio-economic conditions and success.

As part of this study, Statistics Canada representatives will visit 1,100 Canadian schools and meet with 30,000 15-year-old students. They will administer both the international PISA questionnaires and the Canadian YITS questionnaires. They will also interview parents for the YITS and distribute PISA questionnaires to school principals to better understand the factors leading to students’ success or failure in school as well as their transition from school to the job market.

YITS will meet the data needs of the provincial and federal Ministries of Education and Labour who are studying the opportunities that youths will find in the labour market and looking for ways to help them participate fully within society.
Appendix 2. 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.

<table>
<thead>
<tr>
<th>General rules to apply at all times:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Use weighted data: Disclosure can still occur with weighted data but the risk is reduced.</td>
</tr>
<tr>
<td>• Do not report statistics or cells with a fewer than five respondents (or the number reported in the documentation of the survey you are using, whichever is less).</td>
</tr>
<tr>
<td>• No anecdotal information may be given about specific respondents.</td>
</tr>
</tbody>
</table>
The following examples serve as guidelines for dealing with various data types:

**Tabular Output - Frequency Data or Tables of Magnitude:***

<table>
<thead>
<tr>
<th>Data Result</th>
<th>Disclosure Problem</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Low Frequency Cells) Frequency tables or cross-tabulations where a category or cell contains only a few respondents. Or, Estimates from a table of magnitude that has a low frequency cell.</td>
<td>Since the cell applies to only a handful of respondents, they may be identified by knowledge of their characteristics. Small size means five or less or the number provided in the documentation for your survey data.</td>
<td>Collapse classification categories or exclude categories from analysis. In excluding categories, make sure the excluded category counts can’t be derived from marginal totals in the table.</td>
</tr>
<tr>
<td>Reporting a frequency table or cross-tabulation where a category or cell is equal to zero. Or, Reporting an estimate from a table of magnitude where a category or cell has zero respondents.</td>
<td>There are two kinds of zero cells: 1) <strong>structural zero cells</strong> cannot possibly contain a respondent, for example, a cell for ‘married’ and ‘under 12 years old’. 2) <strong>non-structural zero cells</strong> could potentially contain a respondent but do not. Structural zero cells are not a data confidentiality problem. Non-structural, zero cells are a confidentiality problem when they reveal sensitive information about what a respondent is not. For a categorical income variable, the zero cells may present a potential disclosure risk if the remaining (i.e., non-zero) cells represent a narrow range of possible values.</td>
<td>Non-structured zero cells should only be published if they account for less than 15% of the non-marginal cells of a table and if they cause no potential disclosure risk. Otherwise, collapse categories or exclude categories from analysis. For zero cells in ordinal data like income, the highest possible value should not be less than twice the lowest possible value.</td>
</tr>
<tr>
<td>(Full Cell). Reporting frequency or cross-tabulation tables</td>
<td>The data confidentiality risk depends upon the type of information in the table. There</td>
<td>If it has been deemed to be a problem, then change categories, exclude categories.</td>
</tr>
</tbody>
</table>

---

2 Table of magnitude here refers to a cross-classification of a statistic such as Mean wage by gender and full-time/part-time status or Median test scores by gender and grade. Each cell of the table contains a mean or a median and each cell is defined by the values of the classification variables (gender, full-time/part-time status, grade).
<table>
<thead>
<tr>
<th>Data Result</th>
<th>Disclosure Problem</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>where a category or cell contains 100% of the sample (e.g. 100% or the</td>
<td>is little risk in publishing full cells when they reveal something general or easily</td>
<td>from analysis or alternate analysis should be done.</td>
</tr>
<tr>
<td>row or column total in a 2-way table).</td>
<td>known such as the gender of respondents.</td>
<td></td>
</tr>
<tr>
<td>Or,</td>
<td>However, it is more problematic when the full cell reveals sensitive information</td>
<td></td>
</tr>
<tr>
<td>Reporting an estimate from a table of magnitude that has a full cell.</td>
<td>about individuals that would not otherwise be known (e.g., that all 16-17 year olds</td>
<td></td>
</tr>
<tr>
<td></td>
<td>have experimented with drugs).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RDC staff can provide guidance in deciding when a full cell proposes a data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>confidentiality problem.</td>
<td></td>
</tr>
</tbody>
</table>

**Individual Statistics:**

<table>
<thead>
<tr>
<th>Data Result</th>
<th>Disclosure Problem</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporting a ratio.</td>
<td>Ratios pose a disclosure risk if any of the components are deemed to pose a</td>
<td>Do not report a ratio unless its component statistics do not pose a risk of disclosure.</td>
</tr>
<tr>
<td></td>
<td>disclosure risk.</td>
<td></td>
</tr>
<tr>
<td>Reporting a total, mean or average based on less than five respondents.</td>
<td>This is the same problem as reporting an estimate from a table of magnitude with a</td>
<td>Select a more generous sample on which to calculate the statistic.</td>
</tr>
<tr>
<td></td>
<td>small cell size.</td>
<td></td>
</tr>
<tr>
<td>Reporting order statistics such as medians and percentiles where there are</td>
<td>The disclosure risk is similar to that for small cell sizes.</td>
<td>Calculate other order statistics, such as larger percentiles.</td>
</tr>
<tr>
<td>less than five respondents above or less than five respondents below the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>order statistic.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Analytical Outputs:**

<table>
<thead>
<tr>
<th>Data Result</th>
<th>Disclosure Problem</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporting ANOVAs and regression equations.</td>
<td>These analytical outputs do not normally present a disclosure problem.</td>
<td>Be sure that all variables in the model adhere to disclosure rules for</td>
</tr>
<tr>
<td>Data Result</td>
<td>Disclosure Problem</td>
<td>Approach</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Reporting ANOVAs and regression equations when the model involving</td>
<td>These models pose a disclosure problem when the classification groups in the regression</td>
<td>Proceed as when publishing the table whose classification variables are these same covariates,</td>
</tr>
<tr>
<td>categorical covariates is <strong>saturated or nearly saturated</strong> (i.e. has a</td>
<td>have small cell sizes, zero cells or full cells.</td>
<td>and apply the appropriate rules for tabular outputs.</td>
</tr>
<tr>
<td>many coefficients – intercept, main effects and interaction terms – or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nearly as many as there are possible combinations of the covariate values)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reporting scatter-plots, plots of residuals, or box-plots.</td>
<td>These invariably present a disclosure risk because they display values for individual</td>
<td>Graphical outputs should respect all the rules specified elsewhere in this document. It may</td>
</tr>
<tr>
<td></td>
<td>respondents particularly in the case of extreme outliers.</td>
<td>simply not be possible to produce a plot for public release.</td>
</tr>
</tbody>
</table>

**Geography and Indirect Identifiers:**

<table>
<thead>
<tr>
<th>Data Result</th>
<th>Disclosure Problem</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporting the location of a sample cluster on a map, list or otherwise.</td>
<td>Characteristics from the sample might easily be attributed to <em>sampled</em> residents in</td>
<td>Avoid.</td>
</tr>
<tr>
<td></td>
<td>the identified location.</td>
<td></td>
</tr>
<tr>
<td>Reporting tabular outputs on variables such as race or ethnicity below the</td>
<td>This poses a data confidentiality problem particularly when there is a lot of detail for a</td>
<td>Use broad categories such as White/Other, English/French/Other, or Canadian/Immigrant whenever</td>
</tr>
<tr>
<td>national level.</td>
<td>particularly small geographical area.</td>
<td>possible.</td>
</tr>
<tr>
<td>Reporting tabular output for, or by, sub-provincial areas smaller than</td>
<td>The confidentiality problem lies in the increased risk with which an individual in a</td>
<td>Try to avoid output based on such small, easily identified groups. Consult your RDC analyst if</td>
</tr>
<tr>
<td>250,000 people.</td>
<td>small area or highly visible subpopulation can be identified. The problem is the same</td>
<td>this cannot be avoided.</td>
</tr>
<tr>
<td>Or,</td>
<td>as the small cell counts case but often the risk of disclosure is higher in this case.</td>
<td></td>
</tr>
<tr>
<td>Reporting tables that include classification variables that identify very</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Result</td>
<td>Disclosure Problem</td>
<td>Approach</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Or,</td>
<td>Reporting tables that include more than 3 indirect identifiers as classification variables (in addition to the geographical information).</td>
<td>Use only one geographical identifier when possible and avoid reporting tabular data for a single geographic area over time.</td>
</tr>
<tr>
<td>Reporting tables with two or more geographical classification variables (i.e. Health Region, Census Division).</td>
<td>The intersection (i.e. common members) of two or more geographical areas is likely very small (i.e., very few individuals).</td>
<td></td>
</tr>
<tr>
<td>Or,</td>
<td>Reporting on geographical area over time can lead to the identification of persons who have moved.</td>
<td></td>
</tr>
<tr>
<td>Reporting the same geographical classification for two different time periods.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Information about Individual Respondents:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Result</td>
<td>Disclosure Problem</td>
<td>Approach</td>
</tr>
<tr>
<td>Reporting maximum or minimum values for sensitive variables such as income, age, and household size.</td>
<td>Risk of identification is greatest when the maximum or minimum value indicates the presence of an atypical respondent.</td>
<td>Report standard deviations or other statistics that can be used to describe the range of values without reporting an actual maximum or minimum. The case of maximum or minimum in categorical or “dummy” variables may be expected.</td>
</tr>
<tr>
<td><strong>Related Outputs:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Result</td>
<td>Disclosure Problem</td>
<td>Approach</td>
</tr>
<tr>
<td>Reporting similar information from previous studies or cycles of a survey or from other surveys.</td>
<td>This is the residual disclosure problem, i.e., it may be possible to infer individual identities and information by comparing the results to results involving similar sets of classifications (e.g., Two types of geographical classification systems, two different ‘breakdowns’ of</td>
<td>Results involving similar sets of classifications (e.g., Two types of geographical classification systems, two different ‘breakdowns’ of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Result</td>
<td>Disclosure Problem</td>
<td>Approach</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>other released data.</td>
<td>occupational codes, etc.) should be examined closely.</td>
</tr>
</tbody>
</table>

If Public Use Microdata Files (PUMF) are released for the same survey then the published results should not disclose sensitive information about individual respondents that was suppressed from the PUMF. That is, from the reported results, it should not be possible to use data on the PUMF to infer information about an individual that was suppressed in the PUMF.
Appendix 3. On-premises access microdata research contract

**ON-PREMISES ACCESS 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 Contractor),
(hereinafter referred to as the "Contractor"),

and,

(Name and affiliation of Research Assistants),
(hereinafter referred to as the "Researchers"),

WHEREAS Statistics Canada requires the services of the Contractor 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 Contractor and the Researchers are considered ‘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 CONTRACTOR

1. (1) The Contractor 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 Contractor and Researchers are engaged for the sole purpose of providing that service.

ACCESS TO THE MICRODATA
2. Subject to the Contractor and the Researchers undergoing an enhanced reliability check satisfactory to Statistics Canada and taking the oath of secrecy, Statistics Canada shall give the Contractor and the Researchers 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 the Statistics Canada Regional Offices, during normal hours of work, Monday to Friday.

(2) Statistics Canada shall also provide the Contractor and the Researchers with the necessary computing facilities, software and documentation related to the microdata file provided pursuant hereto as is reasonably necessary for the use of the microdata file.

DEPARTMENTAL REPRESENTATIVE

4. The Director of the (Name of Division) is the designated departmental representative for Statistics Canada responsible for the administration of this contract.

LIMITATIONS ON USES OF THE MICRODATA FILE AND PROPOSED OUTPUT

5. (1) The Contractor and the Researchers 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 shall not be used for any other purposes without the prior written consent of Statistics Canada.

(3) The Contractor and the Researchers 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 Contractor and the Researchers 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 unless prior written authorization has been obtained as required by the Record Linkage Policy.

(5) The “Proposed Output” must meet the requirements of both peer and institutional review. Once these requirements are met, the “Proposed Output” and any
information derived therefrom will first be released by Statistics Canada, for example, in one of its flagship publications or in a research paper.

(6) Thereafter, the Contractor may, subject to subsection 6(5), carry out a secondary analysis, but such analysis shall be based solely on the approved “Proposed Output” released pursuant to subsection (5) and be related to the analytical work undertaken to produce the “Proposed Output”.

(7) The Contractor agrees to work with Statistics Canada in trying to meet the requirements of peer and institutional review required for the selected flagship publication/research paper.

(8) In the event the “Proposed Output” fails a peer or institutional review and Statistics Canada decides not to publish the “Proposed Output”, Statistics Canada will give the Contractor written notice to this effect within 30 days of having made the final decision.

(9) Subject to subsections 6(5) and 10(2), in the event Statistics Canada notifies the Contractor in writing that the “Proposed Output” will not be published, the Contractor will not be prevented from:

(1) Publishing the “Proposed Output” elsewhere, and/or

(2) Using the “Proposed Output” for purposes of the attainment of an educational degree.

OWNERSHIP

6. (1) The computer facilities, 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 is a service contract to use 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 computer facilities, microdata file or the related documentation to the Contractor and the Researchers.

(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 fit, the “Proposed Output”, any research, reports, documents, materials or information produced by the Contractor pursuant to this contract.

(3) Copyright in the “Proposed Output”, any reports, documents, materials or information produced by the Contractor pursuant to this contract shall vest in Her Majesty the Queen in Right of Canada. The Contractor 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. Subject to subsection 5(5), Statistics Canada (Her Majesty the Queen in Right of Canada) hereby grants to the Contractor 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 Contractor.

(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 contractor 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.”

SECURITY REQUIREMENTS

7. (1) Any material to be removed from the premises of Statistics Canada by the Contractor or Researchers 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 Contractor or Researchers to take all precautions to avoid disclosure of confidential information or information that may lead to the identification of an individual respondent. The Contractor or Researchers may remove summary files, tabulations and analytical output under the terms of this clause.

(2) The Contractor and the Researchers shall not remove any of the computer materials, original data sets or copies of subsets of the microdata file or any confidential statistical information provided pursuant to this contract from the premises of Statistics Canada.

(3) The Contractor and the Researchers shall be provided with copies of all relevant Statistics Canada policies related to confidentiality, privacy and security and shall acknowledge in writing their compliance with all of the policies.

TERM
8. 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

9. (1) Where the Contractor is in default in carrying out any of its obligations under this Contract, Statistics Canada may, upon giving written notice to the Contractor, terminate the Contract immediately.

(2) The Contract may, by providing 30 days written notice, be terminated by mutual written consent between the Contractor and Statistics Canada.

(3) Any notice to be given to Statistics Canada or the Contractor shall be sent by registered mail to:

(Address of Statistics Canada) (Address of the Contractor)

(4) Any notice shall be deemed to be effective on the day it is received at the address set out above.

PENALTIES

10. (1) As a ‘Deemed Employee’ of Statistics Canada, the Contractor and the Researchers 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 10(1) survives indefinitely the completion of this contract or the termination of this Agreement pursuant to subsections 9(1) or 9(2).

AMENDMENT

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

12. In consideration of the service to be provided by the Contractor, Statistics Canada agrees to pay the Contractor the sum of $_____________.

OR

CONSIDERATION

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

ENTIRE AGREEMENT

13. This contract constitutes the entire agreement between the Parties with respect to the subject matter herein and supersedes all previous negotiations, communications and other agreements unless they are incorporated by reference in this contract.

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

FOR STATISTICS CANADA:

Witness ___________________________ Date ___________________________ Director, (Name of Division)

FOR THE CONTRACTOR:

Witness ___________________________ Date ___________________________ (Name of Contractor)

Witness ___________________________ Date ___________________________ (Name of Research Assistant)

Witness ___________________________ Date ___________________________ (Name of Research Assistant)
Appendix 4. Proposals guideline and layout

RDC application proposals should be short, preferably between two and five double-spaced pages. The format is left to the discretion of researchers, but all proposals are expected to include the following items:

1. Name and affiliation of principal investigator;

2. Names and affiliations of other researchers associated with the proposal;

3. Project title;

4. A concise statement of objectives, including a statement of the research question or questions and an indication of how the proposed research will advance understanding in the area of interest;

5. A statement of the proposed statistical/research methodology, including software requirements;

6. A statement explaining why accessing the confidential data (as opposed to public use microdata files) is necessary. Which confidential data file(s) should be used with a statement explain the data file(s) identified is (are) in fact suitable for the proposed research;

7. An indication of the preferred date on which work on the project would start, and its intended date of completion;

8. Specify which RDC location you wish to conduct the research.

9. Curriculum vitae for each member of the research team (not included in the page limit noted above).

Please note researchers are required to submit a separate proposal for each new research project.
Appendix 4-A. Sample layout of a proposal

**Research Project Title**

**Submitted by**

Name of Division and Director  
Name of Contractor and Researcher Assistants  
Address of Contractor and telephone numbers  
Affiliation

**Proposal**

Describe proposed research project in detail.

**Data Requirements**

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

**Proposed Output**

Indicate the type of output to be generated, such as, a paper for the research paper series or one of the flagship publications or an analytical report with text graphics and tables co-authored by the Contractor and Statistics Canada.

**Completion Date**

Proposed Completion Date: (Enter Dates and expected product)

**Research Location**

Indicate where the researcher is to be located for purposes of this project.

**Source of Funding**

List any agencies that may be providing funds related to this project.
Appendix 5. The Oath

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 all of 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.

I hereby acknowledge having read and understood the following sections of the Statistics Act:

Subsection 17(1) – Prohibition against divulging information

“17. (1) Except for the purpose of communicating information in accordance with any conditions of an agreement made under Section 11 or 12* and except for the purposes of a prosecution under this Act but subject to this section,

(a) no person, other than a person employed or deemed to be employed under this Act, and sworn under section 6, shall be permitted to examine any identifiable individual return made for the purposes of this Act; and

(b) no person who has been sworn under section 6 shall disclose or knowingly cause to be disclosed, by any means, any information obtained under this 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.”

Section 30 – Offences and punishment

“30. Every person who, after taking the oath set out in subsection 6(1),

(a) desert from his duty, or willfully makes any false declaration, statement or return in the performance of his duties,

(b) in the pretended performance of his duties, obtains or seeks to obtain information that the person is not duly authorized to obtain, or

(c) contravenes subsection 17(1)

is guilty of an offence and liable on summary conviction to a fine not exceeding one thousand dollars or to imprisonment for a term not exceeding six months or to both.”

Source: Statistics Act.

*Note: Section 11 refers to the use of the confidential microdata master file. Section 12 refers to the use of the ‘share file’ that is a subset of the master file in which respondents has agreed to share their information with other government agencies, created under the Statistics Act, Section 12. The share file is for the researchers under contract with Human Resource Development Canada.
Appendix 6. Research contract amendment

RESEARCH DATA CENTRES MICRODATA RESEARCH CONTRACT AMENDMENT
(Could be used as an extension of contract, addition of team members)

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 -------------- PROJECT TITLE ------------------;

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

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

Witness Date Co-Investigator
Appendix 7. Sample outline of a Research Paper

Example of what would normally be included in a paper:

Two-page (roughly) Summary of the objectives, approaches and major findings

Introduction
   Issue(s) being addressed
   Literature review

Data sources
   Description of Survey(s) (period, cross-sectional/longitudinal, etc.)
   Information on the Data (i.e. sample size, coverage, exclusions, etc.)
   Data limitations and issues

Methodology/Approach
   Concepts and variables employed
   Model(s) used
   Weighting procedures

Findings
   Interpretation of results from statistical analysis and modeling

Conclusion

Tables and charts
   These may be included in an Appendix as is the standard practice with research papers, or in the body of the paper, or a combination of the two.

Appendices
   An Appendix on some insights into the data including the difficulties encountered and how problems were solved as well as some suggestions for future data collection in the researched area. This would be extremely useful for future researchers as well as for Statistics Canada survey takers and data quality experts.

Bibliography