Research Data Centres Program
Researcher Guide

Microdata Access Division
Statistics Canada

August 2015

Table of Contents

CHAPTER 1: Becoming a Deemed Employee and Accessing the Research Data Centre

To enable Statistics Canada to fulfill its mandate, the _Statistics Act_[1] allows for the use of contractors, researchers, or employees of other government departments to undertake research and analysis. To access protected information to perform these services, such individuals are _deemed to be employed under the Statistics Act._

Research Data Centre (RDC) Researchers must become deemed employees of Statistics Canada[2] to access Statistics Canada confidential microdata.[3] The implications of deemed employee status for RDC Researchers are twofold:

1. RDC Researchers are required to deliver a product[4] (e.g., statistical work, article, thesis/dissertation, etc.) to Statistics Canada upon completion of their project.
2. RDC Researchers are subject to the same policies (and related penalties) under the _Statistics Act_ as are all Statistics Canada employees.[5]

1.0   Deemed Employee Certification Process

The _overall certification process_ of becoming a deemed employee consists of several events that occur _before_ the Director of Microdata Access Division (MAD) signs the Microdata Contract (MRC) granting access to confidential microdata in the RDC:

1. Approval of the proposal
2. Granting of Security Clearance
3. Swearing _or_ Affirming the Statistics Act Oath
4. Acknowledgement of having read relevant documentation: Sections 17(1), 30 and 34 of the _Statistics Act_
5. Acknowledgement in writing (by signing the MRC) having read, understood and agreed to comply with the Researcher Guide, Orientation Session, Values and Ethics Code for the Public Sector, the Policy on Conflict of Interest and Post Employment, and the Statistics Canada Code of Conduct
6. Complete and submit to the Director of MAD, a Confidential Report if there is a potential conflict of interest

Once a Researcher completes this process, (s)he receives the status of a deemed employee. The MRC can then be signed by the Director of MAD, making it a _legally-binding document and_

---

[1] Section 17 of the _Statistics Act_ states: "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."

[2] Reference: Statistics Canada's "Directive on the Use of Deemed Employees".

[3] Individuals who will not be analyzing the confidential data in the RDC, but who come to the RDC as part of a research team and have contact with confidential materials, are also required to follow the same process to become a "deemed employee."

[4] Products are discussed in Chapter 5.3 Submitting a Product to Statistics Canada to fulfil Contractual Obligations

[5] Although being a deemed employee does not create an employment relationship between Statistics Canada and RDC researchers

certifying that the required procedures have been followed. The Researcher can then be granted access to the confidential microdata.

NOTE: <u>Current Access Researchers</u> (those who have continuous access or who have completed a project within the previous year) and <u>Expired Access Researchers</u> (those who have not been active in the RDC for one year or more) starting a new project in the RDC may not be required to complete all the steps of the project certification process again. Please discuss this with an RDC Analyst.

1.1   Security Clearance

A Researcher *must* pass a Government of Canada personnel screening, including a criminal background check before security clearance will be granted. The security clearance is required before a Researcher may be granted to access the confidential data. The personnel screening, including a criminal background check is conducted to establish the person's identity, confirm his/her work/education credentials, and establish his/her personal reliability and ability to be entrusted with Statistics Canada's confidential microdata.

Researchers are expected to complete the Personnel Screening, Consent and Authorization Form for the security check at the RDC in the presence of an Analyst or Statistical Assistant, usually during the orientation session.[6] The Analyst witnesses the signing of the form and confirms the Researcher's identification. Fingerprints may also be requested by the RCMP as part of the security clearance.[7]

1.2   Signing a Microdata Research Contract (MRC)

As stated in Section 4.4 of the MRC, by signing the MRC Researchers are acknowledging that they will comply with the following:

- The Research Data Centre Researcher Guide
- The Research Data Centre Orientation Session
- The Statistics Act Oath or Affirmation of Office and Secrecy
- The Values and Ethics for the Public Sector
- The Policy on Post-Employment and Conflict of Interest
- The Statistics Canada Code of Conduct

The Contract specifies the following terms of access:

- Agreement that the work to be done in the RDC by the Researchers and the results produced are to correspond to the objectives identified in the research proposal[8]
- Agreement by Statistics Canada to provide access to specified confidential microdata[9]
- Agreement of the Researchers to abide by the RDC security and confidentiality requirements

---

[6] For out-of-town researchers, please contact the RDC analyst to discuss alternative options
[7] Please consult an RDC Analyst for detailed information about a request for fingerprints.
[8] A Researcher is to submit a new proposal for any subsequent research to be carried out at an RDC. In addition, SSHRC and/or Statistics Canada may ask for a new proposal if the scope of the research changes significantly during the course of the project. Please consult the RDC Analyst for more information.
[9] Please refer to section 1.5.2 for information about adding external data to a research project.

- Agreement on the length of the contract
- Agreement of the Principal Investigator to deliver a product[10] to Statistics Canada

> **Only RDC Researchers who are *listed on a project's MRC and have completed the Deemed Employee Certification Process* may examine the confidential data files for a project**

1.3   Attending an Orientation Session

A new Researcher[11] will be invited to attend an RDC orientation session. The orientation session is designed to familiarize Researchers with the responsibilities associated with Statistics Canada deemed employee status and the security measures surrounding access to the RDC. During the orientation, many of the steps of the certification process will be completed.

1.3.1   The Statistics Act Oath or Affirmation of Office and Secrecy

The Statistics Act Oath or Affirmation of Office and Secrecy (Statistics Act Oath) must be sworn or affirmed aloud by the Researcher prior to being given access to the confidential microdata. The Oath, affirmed by all employees and "deemed employees" of Statistics Canada, conveys the obligation of the deemed employee to protect the confidentiality of the data.

The Statistics Act Oath states that each employee or deemed employee of Statistics Canada will:

> **"........solemnly swear (or affirm) that I will faithfully and honestly fulfill my duties as an employee of Statistics Canada in conformity with the requirements of the Statistics Act, and of all 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."**

The Oath is legally binding *for life*;[12] hence even after Researchers have completed their research contract, they are not permitted to reveal any confidential information. Violations, as identified by the Chief Statistician, are punishable by a lifetime ban from access to Statistics Canada detailed microdata. As well, they are subject to prosecution and are liable on summary conviction to a fine and/or imprisonment.[13]

> **The *Oath of Office and Secrecy* requires Researchers to be <u>personally</u> accountable to uphold the confidentiality provisions of the Statistics Act and not to reveal anything about individual respondents, either directly or indirectly.**

---

[10] Information about products can be found in Chapter 5.3 Submitting a Product to Statistics Canada to fulfil the terms of the Microdata Contract (MRC).

[11] For Researchers who have been out of the RDC for more than year, it is highly recommended that the orientation session be given again. This is an opportunity to review the policies and procedures of the RDC program, some of which may have changed.

[12] Although binding for life, Researchers may be requested to reaffirm the Oath.

[13] Please refer to the *Statistics Act*, Section *30.* See also Statistics Canada's "Directive on the Security of Sensitive Statistical Information", Appendix B.

1.3.2 Values and Ethics Code for the Public Sector & the Policy on Conflict of Interest and Post-Employment

After a Researcher has sworn or affirmed the Statistics Act Oath and becomes a deemed employee of Statistics Canada, (s)he agrees to follow all rules and instructions under the Statistics Act, as well as Statistics Canada policies. Two important documents outlining Statistics Canada policies are the *Values and Ethics Code for the Public Sector* and the *Policy on Conflict of Interest and Post-Employment*.

Researchers, as deemed employees of Statistics Canada using the RDC facilities, must conduct themselves in accordance with the principles and spirit of the *Values and Ethics Code for the Public Sector* to prevent real, apparent or potential conflicts of interest from arising. A conflict of interest is a situation in which a public servant or deemed employee has private interests that could improperly influence the performance of his or her official duties and responsibilities or in which a public servant or deemed employee uses his or her office for personal gain. A *real* conflict of interest exists at *present*. An *apparent* conflict of interest could be *perceived* by a reasonable observer to exist, whether or not it is the case. A *potential* conflict of interest could reasonably be foreseen to exist in the *future*. The *Policy on Conflict of Interest and Post-Employment* elaborates on the Values and Ethics Code for the Public Sector.

Each RDC Researcher, as a deemed employee, reads the *Code* and *Policy* to interpret his/her particular situation for *each* contract that (s)he signs. The researcher should keep three issues in mind to determine whether (s)he has a conflict of interest:

1. The research will be used only for the purposes indicated in the proposal
2. All Researchers involved in the project have completely revealed all their sources of funding and support[14]
3. No one (or organization) outside of those persons (or organizations) listed in the proposal has an 'interest' in the project

If the Researcher cannot completely agree with all three of these statements *prior to or during the life of the RDC contract*, or if there are any other clauses in the *Code* or *Policy* that may be an issue for the researcher, it is *his/her responsibility* to discuss the situation with the Analyst and/or Director of Microdata Access Division and, if applicable, submit a *Confidential Report* to Statistics Canada.[15]

---

[14] If sources of funding change during the life course of the project, researchers are obligated to report this to the Analyst / Director of Microdata Access Division

[15] Deemed employees must, within 60 days after appointment, or following a change during the appointment, make a Confidential Report to the designated official of all activities that might give rise to a conflict of interest in respect of the deemed employee's duties and responsibilities. Deemed employees should not complete this form if there are no activities that give rise to a conflict of interest.

1.3.3   Statistics Canada Code of Conduct

The *Statistics Canada Code of Conduct* integrates the fundamental values and commitments of the [Values and Ethics Code for the Public Sector](#) with specific expectations for the behaviour of all Statistics Canada employees, including deemed employees. Of particular relevance to the RDC are:

The core value of accountability
- Employees are required at all times to protect the confidentiality of collected data and to uphold the public trust in Statistics Canada
- Confidential information must be shared on a *need-to-know basis*—only with appropriate personnel

> **RDC Researchers may only discuss confidential microdata with Statistics Canada employees or other deemed employees listed on their own MRC**

The expected behavior of compliance with the Network-Use Policy
Researchers must not conduct any unlawful or unacceptable activity, such as:
- Attempting to defeat information technology security features, through such means as using anti-security programs; using someone else's password, user-identification or computer account; disclosing one's password, network configuration information or access codes to others; or disabling anti-virus programs
- Destroying, altering or encrypting data without authorization and with the intent of making it inaccessible to others with a lawful need to access it

1.4   Access to RDC Data

Once all the steps in the certification process outlined above are complete, and the after the MRC has been signed by the Director of MAD, a Researcher can then be given access to confidential microdata inside the RDC.

A Researcher *RDC user account*[16] will provide access to:
- A secure survey drive with the data file(s) and relevant documentation, such as codebooks and user guides.
- A project drive where each project will have an individual project folder. Within their folder, Researchers can create and store syntax, data and other files.

> **An RDC Researcher's access is limited to the data sets listed on his/her MRC.**

1.5   Use of Data in Research Data Centres

The RDC program recognises that in some cases, combining data from different sources has the potential to strengthen research by increasing sample size, providing contextual

---

[16] Please refer to Section *3.2 Computer Security in the RDC* for more detail about RDC user accounts.

information, and so on. This section discusses some of the things to know about the use of data from sources other than RDC microdata files in the RDCs.

1.5.1   Publically Available Statistics Canada Aggregate Data

Aggregate data produced from Statistics Canada data found in the public domain do not require special approval to be used in the RDC. Some examples include:

- The Consumer Price Index
- Census profiles
- Employment related rates
- CANSIM tables

Researchers are to direct their requests to their Analyst by providing the Analyst the aggregate data to be added to their project folder. The Analyst may request to verify the source of the data before the data can be loaded.

1.5.2   Administrative Data (Formerly Non-Statistics Canada Data)

For Researchers to be able to add Administrative data (non-Statistics Canada[17] data) to their RDC projects for use with Statistics Canada microdata files, the following conditions must be met:
1. A clear description of the external data file (either in the project proposal or in an amendment to the proposal), the type and source of the data (including the unit of analysis), and how the data are going to be used in the analysis.
2. The researcher needs to demonstrate in writing that (s)he has permission from the source of the data to use these data in the RDCs. Evidence of permission might include a letter of agreement from the data holder.[18]
3. If the use of the data requires a record linkage the researcher will work with the appropriate Subject Matter Division and the Chief of the RDC program to obtain Statistics Canada's Executive Management Board approval before the data can be brought into the RDC.[19]

When the Researcher brings in an approved external dataset for use inside the RDC, the data would be available only for the team listed on the MRC for that particular contract. Once the MRC has ended,[20] the data will not be returned to the Researcher.

1.5.3   Use of Share Files in Research Data Centre

Government departments who are funding partners to a survey may have contracts with Researchers to conduct research on their behalf. When there is an external funding source to a Statistics Canada survey, each respondent is asked if (s)he agrees to share his/her information with the funding government department.

---

[17] The Postal Code Conversion File (PCCF) is no longer considered a Statistics Canada product. For a Researcher to use the PCCF in an RDC, the file must be listed on the MRC or added as an amendment to the contract.

[18] Please consult an RDC Analyst for details.

[19] Please consult an RDC Analyst for details.

[20] Please refer to the section on archiving in Chapter 5 for information on the preservation of researchers' working files, which may contain elements of the external data.

A share file is a subset of the master file containing only the data from those respondents who agreed to share their information with the partner agency. All other data from respondents whom did not agree to share their responses are removed from the share file. By contrast *a master file contains the data from all survey respondents*.

Researcher responsibilities and the use of Share files in the RDCs

Researchers accessing a share file in an RDC have a responsibility to protect the share file data. A Researcher may have other projects in the RDC in which he or she has access to the master file of the same data at the time a share file project is approved.

- Researchers *cannot* use information from the master file to inform the share file.
- When conducting research with share files, statistical information from the corresponding master file in any form (verbal, written, electronic or printed) cannot be used to support the research in a share file project.

1.5.4   The Use of Tax Data in the RDC

Additional confidentiality documentation must be read and understood, by all Researchers who have access to RDC data files which include tax data. This includes any survey that asked respondents permission to link their tax information to their survey responses in place of respondents self-reporting their income. This applies to most of the Statistics Canada Household Surveys. [21]

Researchers must read and understand Sections 239 and 241 of the *Income Tax Act* and Sections 295 and 328 of the *Excise Tax Act*, if the statistical information to be accessed under this Contract includes taxpayer information obtained from the Canada Revenue Agency. These sections include provisions of data access and offences and punishments for the misuse of tax data.

---

[21] To find out whether data on the MRC include tax data, please consult the survey documentation.

CHAPTER 2: Contract Management

During the life course of a contract, changes inevitably occur in the team members or the data required. This section provides an explanation of what changes can occur and the procedures in place to manage those changes in a contract's life cycle. This information mostly applies to academic researchers with SSHRC approved proposals. Most of the material on contract amendments also applies to government funded research project or [Programs of Research](.)[22]

2.1   Contract Amendments

The Principal Investigator is to uphold the contractual obligations on behalf of the team and has administrative control of the MRC. This means that only the Principal Investigator can request amendments to the MRC such as adding or removing Co-Investigators or datasets, requesting extensions and so on.

2.1.1   Add/Remove a Co-Investigator

The Principal Investigator may request to have a team member added or removed[23] from a contract. The Principal Investigator provides the Analyst with a written request to make such changes. When adding a Co-Investigator, who is not already current deemed employee the new team member will go through the deemed employee certification process.[24]   An amendment to the contract is signed by the Principal Investigator and the Co-Investigator being added or removed.[25]

2.1.2   Add a Dataset

Researchers are granted access only to the microdata requested in their approved proposal. On occasion, it may be appropriate to add a data set to the original contract in order to answer the research question defined in the approved proposal. An example of an appropriate case would be when the sample size in the original data set is not of sufficient size to answer the original research question and another data set is available with similar information and a larger sample size.

The Principal Investigator submits a written request to the Analyst detailing the rationale for why a new data set needs to be added and outlines a commitment to maintain the objectives of the original proposal. Addition of the microdata[26] to the MRC is contingent upon a positive institutional review by Statistics Canada.

---

[22] Please consult an RDC Analyst for details.

[23] The removed Co-Investigator will no longer be permitted access to the RDC to work on the Principal Investigator's contract.

[24] Please see an RDC Analyst for more information about the process of adding or removing a team member.

[25] If there are extenuating circumstances regarding contacting a Co-Investigator for removal from a contract, please see an RDC Analyst.

[26] It is also possible to add non-Statistics Canada data to the MRC to inform the analysis with contextual information (see section 1.5). Please contact the RDC analyst for required documentation to amend this data to the MRC. Use of this data is contingent upon the approval of the Director of Microdata Access Division.

> **The scope of the research question should not change with the request to add data.**

2.1.3   Request an Extension for a Contract

If the team realizes they have underestimated the time needed to complete the research defined in the contract or other extenuating factors have prevented completion of the contract by the end date, the Principal Investigator can request a contract extension.

Contract extensions are ideally negotiated before the contract expires and are contingent upon the approval of the Director of MAD. The amendment to extend the contract is signed by the Principal Investigator (on behalf of the team). Typically, with good reason[27], a contract can be approved for an extension of six months to a maximum of one year past the contract end date.

The Principal Investigator should provide the Analyst with a written justification for the extension. Valid reasons for a contract extension include: [28]
- The data analysis is in the final stages and the extension will cover the completion of the data analysis.
- Funding or staffing shortages have prevented progress on the contract to date (including leaves of absence, long-term illnesses, etc.).
- There have been complications in obtaining data to complete the analysis.

> **In all cases, when requesting to extend a contract, the scope of the research questions must not change. [29]**

2.2   Files Transfers and Contract Transfers

If the Investigators of a project are not located in the same area and require access to more than one RDC, the primary RDC is usually the RDC of the Principal Investigator. The other RDCs are called secondary RDCs. Shared folders may be set-up so Co-Investigators can share project files between centres. The management of the contract is done by the Analyst in the primary RDC. For *permanent relocation of Investigators*, a Principal Investigator can ask the Analyst to transfer his/her contract along with the working files to another RDC. [30]

---

[27] In some rare cases, a dormancy period may be appropriate where no one listed on the MRC has accessed the RDC for an extended period. Please consult with an RDC Analyst for more information.

[28] If the researcher(s) are unable to work on the project for an extended period of time during the life of the contract (for example due to maternity leave or health reasons) the Principal Investigator may request in writing that the contract be made "dormant" for a specified period of time. After the dormancy period, the contract is either given a new end date or the Principal Investigator can withdraw it. Please discuss with an RDC analyst.

[29] If the scope of the research changes, a new proposal should be submitted.

[30] Please contact an RDC analyst for more information.

2.3   Sub-Contracts[31]

On rare occasions, when the proposed research questions remain the same but the purpose of the project has expanded, it may make sense to consider a sub-contract rather than have the researcher apply for a new contract. For example, a student Co-Investigator wishes to expand the on-going project in order to carry out research for his/her dissertation. A separate contract with the student as Principal Investigator is preferable since the student may wish to bring in different team members, employ different statistical methods, or have different time lines. The final product for this type of sub-contract would be the thesis work.[32]

2.4   Contract Withdrawals

It is possible for the Principal Investigator to withdraw his/her contract by providing a written request to the Analyst. Valid reasons for withdrawing a contract include, but are not limited to, insufficient sample size, change in staff or funding for the project, or new research priorities. Requests to withdraw a project will be accepted if there has not been a significant amount of output released for the specified project. Otherwise the Principal Investigator may be asked to provide a summary report briefly explaining the analysis, the outcome of the analysis, and the reason for discontinuing the study.

2.5   Revision contracts

After the contract end date, a Researcher no longer has access to his/her research or project for that expired contract. However, a Researcher may apply for access to a closed project by means of a "revision contract". *Revision contracts are not an extension of the original MRC but are new contracts with new terms of data access* (i.e., new start and end dates and new product obligations).

A revision contract may be requested if the original MRC has been completed but a Researcher needs to re-access RDC project files for one of the following reasons: [33]
1. Revise and resubmit request from an Academic Journal Editor
2. Student revision to a Thesis / Dissertation following a Defense

The duration of a revision contract is typically six months and the contract is for the express purpose of revising a product in order to achieve approval for publication, not for undertaking   new   research   hypotheses.

---

[31] This does not refer to sub-contracts for Programs of Research. Please consult an RDC Analyst for more information about Sub-contracts for Programs of Research.
[32] Please consult with an RDC Analyst for more information.
[33] Researchers will be requested to include Reviewer comments with the Revision Contract request.

CHAPTER 3: Culture of Confidentiality

Promoting a culture of confidentiality and maintaining the confidentiality of survey respondents is of paramount importance to Statistics Canada. It is the trust of the respondents that makes it possible for Statistics Canada to provide valuable data on the socio-economic condition of Canadian society. Under the *Statistics Act*, all employees and deemed employees must protect data confidentiality.

In the RDC, there are additional important elements of the culture of confidentiality. These include the physical and computer protection of the RDC, as well as the confidentiality vetting of research outputs before the results are released from the RDC.[34] This chapter focuses on *physical security* of the RDC and *computer protection*.

3.1   Physical Security of the RDC

The RDC is a physically secure research facility, with reinforced ceilings, floors and walls and the door is either solid core or steel with tamperproof hardware and magnetic locks. Each centre also records RDC entries through electronic access cards.

Access to the RDC is contingent upon having:

a) A personnel security clearance
b) An active Microdata Research Contract (MRC)
c) A *swipe card* to open the door to the RDC
   - This pass must not be lent to or shared with others
   - Researchers are *not permitted* to let other people (including other RDC Researchers) into the RDC
   - If a security pass is lost or stolen, please report it immediately to RDC staff

> **Visitors are _not_ allowed in the RDC, unless they are permitted entry and accompanied by RDC Staff**

3.2   Computer Security in the RDC[35]

Researchers play a crucial role in computer security through responsible use of information technology in the RDC.

Secure Network

- The RDC has a secure network environment with layers of computer security. Each workstation is connected to a central server via a secure wide area network with no

---

[34] see Chapter 4
[35] Reference: Statistics Canada's IT Security Policy.

connectivity to the Internet. This means there is no e-mail or Internet access from any workstation.

- Researchers log off the workstation after every work session. If a Researcher is away from the workstation during his/her work session, the researcher locks the computer to prevent others from accessing it.

User Accounts

- Each research team has access to password-protected common folders on the RDC server where all materials generated for the project are stored.
- The password is <u>not</u> to be shared.
- **Only Researchers listed on the MRC are permitted to view / discuss the approved data files for a project.**

Drive Connectivity & File Transfers

- Workstations have the drives/connections for all storage devices (e.g., USB flash drives, external hard drives, CDs, etc.) disabled.
- Electronic file transfers are conducted through the Analyst.
- Files entering or exiting the RDC will be reviewed by the Analyst before being loaded to or extracted from the RDC server.
- A Researcher cannot install software on an RDC computer.

Personal Responsibility & Electronic Devices in the RDC

- Although Researchers are permitted to carry their personal electronic devices with them in the RDC, they should be out of view when the researcher is in the vicinity of (secure) computer workstations or near any confidential material (e.g., printed outputs).
- Researchers are *prohibited* from operating any communication, text editing/messaging, or image/text capture devices (e.g., laptop computers, tablets, E-readers, cellular phones or other devices with photographic capability) in the vicinity of (secure) computer workstations or near of any confidential material.
- A cell phone (or similar communication device) must never be used to communicate/discuss confidential data or outputs with team members.
- To answer a call, the researcher can do so away from the (secure) computer workstation by conducting the conversation outside of the RDC secure area.

If Researchers wish to have a conference call with team members in another RDC, it can be arranged through the Analyst where facilities are available.

Discussing confidential material on a wireless device is a <u>breach</u> of security.

Summary of Key Principles of *Researcher Responsibilities* for Physical and Computer Security		
Elements	*Controls*	*Responsibilities of the Deemed Employee*
Physical Security	Controlled Entry- Swipe Cards	• Only *deemed employees* are permitted to enter the RDC using their **own** swipe card. • Researchers are not to share their swipe card with anyone else. • Researchers many not let anyone into / out of the RDC.
Computer Security	Access to Data / Project files	• Researchers are **ONLY** permitted to view data files which are listed on their active MRC. • **ONLY** the Principal Investigator and Co-Investigators listed on the approved Microdata Contract are permitted to access / view the approved project files and data files. • **Although working in close proximity to other RDC Researchers, Researchers are not permitted to look at or discuss data and / or results with Researchers not listed on their active MRCs.**
	Procedures prohibiting use of electronic devices	• Researchers are **prohibited** from operating any communication, text editing/messaging, or image/text capture devices *(e.g., laptop computers, tablets, E-readers, cellular phones or other devices with photographic capability)* in the vicinity of (secure) computer workstations or near of any confidential material.

CHAPTER 4: Confidentiality Vetting and Removing Output from the RDC

Another component of the RDC culture of confidentiality is the confidentiality vetting of research outputs before they are released from the RDC. *Confidentiality vetting is the process of screening research outputs, syntax or any other material stored inside the RDC (e.g. written notes or outputs on the RDC server)*, as per Statistics Canada's *Directive on* the *Security of Sensitive Statistical Information*. Vetting occurs when a Researcher wants to remove material from the secure area of the RDC. Once the files are released, the research results are considered public.

4.1   Managing Research Outputs Leaving the RDC

Microdata, in any form, never leave the secure area of the RDC. This also applies to any approved external microdata or aggregate data brought in by Researchers to be used in conjunction with the confidential data housed in the RDC.

Output produced, used and stored *inside* the RDC is not subject to confidentiality vetting. The vetting rules are applied once a Researcher requests to remove output from the RDC.

All outputs and syntax (hand-written, printed or any files stored on the RDC server), leaving the RDC must go through confidentiality vetting; the researcher should screen the results first and then submit them to the Analyst for confidentiality vetting.

> **Researchers must comply with Statistics Canada rules and regulations on data confidentiality when removing output and material from the RDC.**
>
> **The removal or attempted removal of confidential information from the RDC may result in a ban from accessing RDCs and all Statistics Canada confidential microdata.**

4.1.1   Preparing Output for Confidentiality Vetting

When a request is made for output to be released (such as results, syntax or notes), the researcher completes a Vetting Request Form. This provides the required information for the Analyst to conduct the vetting request. Please schedule a time with an RDC Analyst before the first vetting request to review the form.

Once it has been determined that the requested output complies with the confidentiality vetting rules, the vetted files can then be released to the researcher.

Researchers are expected to request outputs based on *need* (such as for conference presentations or drafting an article for publication); this helps to reduce the risk of residual disclosure[36] and facilitates the timeliness of the vetting process. While Analysts will attempt

---

[36] Residual disclosure is the disclosure of confidential information by the combination of released outputs, perhaps with data from external sources. For example, a "residual" or "shadow" table is produced by subtracting sub-sample tabular output from full sample tabular output.

to process a confidentiality vetting request in a timely manner, Researchers are to account for the processing time when planning research activities and submitting output for release.

Non-releasable outputs

Here are some examples of outputs that are not releasable from an RDC:

- Unvetted results printed on coloured paper in the RDC
- Output based on less than the minimum unweighted cell count[37]
- Residual tables with low cell counts
- Anecdotal information
- Data, in any form
- Output that reveals details about the sample frame (i.e., list of postal codes)

Intermediate Output

During the initial stages of a project, many descriptive and tabular outputs may be generated that are intermediate and are used to determine subsample and final variables. Requesting these outputs for release may create difficulties in the release of future output utilizing these subsamples or variables.[38] **It is recommended that each team member who needs to review intermediate output has access to the RDC for the purposes of meeting and reviewing output in the RDC prior to requesting its release**.

4.1.2   Weighting and confidentiality

In general, descriptive output from the RDC based on *sample* data must be weighted.[39] For any descriptive results, Researchers are asked to produce the same outputs both weighted (by applying the survey weights) and unweighted (outputs from the raw data).[40] The weighted file will be for release and the unweighted file will be supporting documentation to ensure confidentiality protection and will not be released.

Statistics Canada household surveys are based on complex sample designs that include stratification, multiple stages of selection, and unequal probabilities of selection. Even when population estimates are not the main interest of the research, weighting corrects for sample bias arising from the survey design. Sample bias can arise from over/under sampling population sub-groups, response bias, data collection and processing operations.[41]

---

[37] Please consult an RDC Analyst for the minimum unweighted cell counts for the survey(s) listed on a MRC.

[38] Please consult an RDC Analyst for more information about limiting requests for intermediate descriptive output and the risk of residual disclosure.

[39] In very limited circumstances unweighted results can be released. Please discuss the vetting rules and the use of weights for the data with an RDC Analyst.

[40] Where survey weights are available. When survey weights are not available (ie., Administrative data files), different rules will be applied to protect confidentiality. Please see an RDC Analyst for details.

[41] Researchers are expected to read all pertinent documentation on weighting with regard to dissemination guidelines as specified in the User Guide for the survey. Dissemination guidelines, which focus mainly on the quality of tables, are not the same as confidentiality vetting rules and RDC Analysts will not be vetting the quality of outputs. Consult with an Analyst for more information.

Weighting plays a role in data confidentiality because one weighted observation may no longer represent one sample respondent.

- Observations in a survey are a sample, drawn from the population of interest.
- When the survey weights are applied, each respondent in the survey represents a certain number of persons in the target population.
- Weighting makes results from a sample generalizable to the population.
- When descriptive statistics and tabulations have been weighted, the information is less likely to be attributed to a specific individual. The weighted frequency is no longer a count of the respondents in the sample.

4.1.3   Responsibilities of Researchers

The most important element of protecting respondents' confidential information is Researchers who exercise caution when removing vetted outputs from the RDC and when reporting and publishing results. Researchers are responsible for applying the vetting rules to the outputs that they are preparing for release.

> **Researchers are *personally* responsible for preventing the disclosure of confidential information.**

Types of Output that vetting rules focus mainly on are:
- Frequency data
- Tables of magnitude (e.g., averages and totals)
- Individual statistics (e.g., minimum and maximums)
- Models which are equivalent to tables (e.g., a model with a single categorical variable or a model with all possible interactions)
- Low cell counts and Residual tables with low cell counts
- Anecdotal information
- Narrowly defined geographical areas, or other sensitive variables[42]

Considerations when preparing output for vetting:
- Provide weighted and unweighted outputs for all counts, frequencies and descriptive statistics
- Complete the vetting request form
- Limit the number of requests
- Consider processing time when submitting request
- Save all syntax files

> **Researchers are responsible for properly applying all vetting rules <u>before</u> submitting output to the Analyst for screening.**

The rules and regulations also apply to hand-written notes produced in the RDC. At the end of a work session, hand-written notes are to be given to the Analyst for shredding, storage, or vetting. If Researchers wish to leave the RDC with their hand-written notes, those notes are to

---

[42] Check with the Analyst to see if the output contains sensitive variables

be checked by the Analyst only (Statistical Assistants are not authorized to vet notes). If the Analyst determines the notes contain no confidential information, the Researcher can leave the RDC with them.

Summary of Key Principles of *Researcher Responsibilities* for Protecting the Confidentiality of <u>UNVETTED</u> Materials		
Elements	*Controls*	*Responsibilities of the Deemed Employee*
Confidentiality Protection	Procedures for protecting Handwritten notes	• Researchers are encouraged to make electronic notes rather than handwritten notes. • If a Researcher makes handwritten notes, the notes should be (a) stored inside the RDC in a locked cabinet or other approved means of storage, or (b) be submitted to an Analyst for shredding, or (c) be submitted to an **Analyst** for approval. If approved by the Analyst, the notes may leave the RDC.
	Protection of Output Printed on Coloured Paper[43]	• Researchers are discouraged from printing output • Where printing on coloured paper is permitted, output must be properly stored inside the RDC or shredded before leaving the RDC.
	Rules around verbal discussions of unvetted confidential information	• Researchers are not permitted to discuss unvetted results with **ANYONE** (including co-investigators on the contract) **outside the RDC**. This includes over the phone when a Researcher is in the RDC and the co-investigator is outside the RDC. • Researchers are **not permitted** to discuss unvetted results *inside* the RDC with Researchers whom are *NOT* listed on their active MRC.

4.1.4   Responsibilities of the Analyst

The Analyst's primary responsibility in confidentiality vetting is to ensure confidentiality is not breached when allowing research outputs to leave the RDC. The Analyst reviews all the materials that the researcher would like to remove from the RDC. The final responsibility and decision to release the output rests with the Analyst. In the event a potential confidentiality issue is identified, all attempts are made by the researcher, in collaboration with the Analyst, to find a solution that protects the confidentiality of the data.

---

[43] The colour of confidential paper may vary by RDC. See an Analyst for details.

4.2 Best Practices for Confidentiality Vetting Requests

Here are some ways that Researchers can minimize delays in the vetting process:

Do
✓   Schedule time with the Analyst well in advance of an initial vetting request and any anticipated deadlines. Check that the confidentiality vetting rules are understood and that the supporting documentation requirements are clear.
✓   For descriptive statistic outputs, provide weighted and corresponding unweighted output in separate and well labeled files.
✓   Restrict crosstabular analysis to two or three dimensions.
✓   Limit the release of tabular output to just before the end of the project.
✓   Take care when using small subgroups or small areas. Remember that the confidentiality vetting rules become stricter for sensitive variables, including detailed levels of geography, and that release of output is sometimes prohibited regardless of the minimum cell count and weighting requirements have been met.
✓   Check with the Analyst for specific rounding rules on descriptive statistical outputs that are applicable to some surveys and administrative data.
✓   Atypical outputs may require extra supporting documents. Check with the Analyst.
✓   Check for errors and remove unneeded results before submitting files for vetting.
✓   Provide all necessary details in the Vetting Request Form. Inform the Analyst if similar output has been previously released (include the details of the differences between similar requests). If applicable, separate requests with differently defined working samples and provide the Analyst the corresponding sample sizes.
✓   Check survey documentation dissemination guidelines which will discuss data quality issues (such as sampling errors, non-sampling errors, coverage errors, response errors, non-response errors, and processing errors).[44]
Do not
✗   Request tables for which the corresponding un-weighted cell counts are below the minimum requirement.
✗   Request anecdotal information (such as min/max statistics).
✗   Request graphs with individual data points plotted or show outliers (e.g. scatter plot of the raw data or box-plots).
✗   Request unweighted outputs without first discussing this with the Analyst.
✗   Discuss unvetted outputs or data with other RDC Researchers or persons whom are not listed on the MRC.

---

[44] While dissemination guidelines may seem similar to vetting rules, they are not the same. Both the dissemination guidelines and the vetting rules should be consulted, but for different reasons. Please note: Analysts will screen output using vetting rules, not dissemination guidelines.

CHAPTER 5: The Completion of the RDC Project

When the contract is close to its end date, the Principal Investigator is responsible for closing out the project.

5.1   End of Contract Letter and Exit Survey

Approximately two months prior to the end of the contract, the Principal Investigator will receive and End of Contract Letter from the RDC Program. The End of Contract Letter will assist Principal Investigators in closing a contract, extending a contract, or revisiting a contract at a later date for future work.

The letter is accompanied by link to a short *Exit Survey*, which gathers information about the researcher's experiences with the RDC program. The Principal Investigator shares the link to this survey with all project team members who have accessed data in the RDC.

5.2   Archiving, Period of Retention and Disposal of Files

It is the Researcher's responsibility to delete any copies of data files or sub-datasets that were not essential to the analyses to free up space on the RDC server. All output released is stored, as are any essential programs or syntax, as requested by the researcher *for 10 years following the contract end date.*

In compliance with Statistics Canada's directives on the retention and disposal of microdata files and aggregate statistics, the files created by RDC Researchers will be kept for a period of 10 years after the completion of the project. The chart below describes the key dates and activities associated with the RDC process for archiving, retention and disposal of project files.

Key Dates & Activities for Archiving, Retention and Disposal of RDC Project Files		
Dates	*Key Activities*	*Explanation*
End of Contract	Request Syntax	• Researchers are advised to request the release of syntax files since Researchers cannot return to the RDC to examine material after the contract end date.[45]
	Request Vetted Output	• Researchers are advised to request the re-release of any previously released outputs that they may want for their records before the contract end date.
9 Years after contract end date	Reminder of Retention and Disposal Directives	• RDC Program will attempt to contact the research team 9 years after the end of the project to remind them about the retention and disposal directives. It will be possible to request a copy of vetted material.
10 Years after contract end date	Deletion of all project files	• ***Ten years after the end of the contract, <u>all</u> files from that project will be deleted.***

---

[45] Access to the RDC will end with the contract and researchers are to take with them all the materials that they brought into and stored in the centre. Confidential material that cannot leave the RDC (such as output printed on coloured paper) will be destroyed. If applicable, access cards are returned to the Analyst.

5.3 Submitting a Product to Statistics Canada to fulfil Contractual Obligations

The contract is complete when a product is submitted to Statistics Canada and thus the contractual obligations are fulfilled. An electronic copy[46] of the product should be submitted to the Analyst on or before the contract end date[47]. The expected product is described in *Appendix C (i.e., "proposed output")* of the MRC. Each contract will require *one of the four* following types of products to be specified in Appendix C:

Product Types: Examples of products that meet the requirements of MRCs are given for each type :[48]
1. Analytic research paper that Statistics Canada may use for internal purposes, or may publish itself.   • *Example products include: Journal article, dissertation, thesis or similar.*
2. Written data quality assessment of frames, or of protected information.   • *Example products include: A technical research paper that describes a data or methodological problem discovered during the course of research (typically chosen when a product cannot be achieved).*
3. Non-protected outputs that are confirmed by a senior governmental official as being important to the work of that official's government department.   • *Example products: Vetted output selected by the government department (for use with sensitive analysis on government projects only).*
4. Specific work related to a particular statistical program which has been specified in advance as part of a legal agreement.   • *Example products include: Vetted output stored in the project file in the RDC. (At the MRC end date, the contract will be complete and no further product is required).*

---

[46] If applicable.

[47] If a product is not received, the Principal Investigator may receive a delinquent letter and until a product is received, the Principal Investigator may be prevented from participating in any future proposals as the principal investigator (PI) and/or as a co-investigator.

[48] For details about what products are acceptable please consult an RDC Analyst.

CHAPTER 6: Publications, Dissemination and Follow-up

This chapter includes information RDC Researchers need to know for publishing outputs generated from Statistics Canada microdata, as well as information about sharing publication information with the RDC Program and the [Canadian Research Data Centre Network](#) (CRDCN).

6.1　Sharing Publication Information is Critical:　Annual Researcher Survey

Every year, the RDC Program and the Canadian Research Data Centre Network conduct a survey of RDC Researchers, which collects and catalogues publications resulting from RDC research. *All RDC Researchers* will be sent a request to complete the Annual Researcher Survey online. The importance of this survey is twofold: to increase the visibility of research stemming from work in the RDCs, and to enable the RDCs and the CRDCN to provide reports as required by agencies which fund the RDCs.　Therefore filling out the survey helps the RDC maintain funding.

By completing the survey, Researchers will increase the visibility of their research.　For example, citations for publications will be listed in the online [CRDCN bibliography](#). Publications could also be further highlighted by being featured in a CRDCN Synthesis Series, in CRDCN Research Highlights, or in a webinar. Please visit the [CRDCN's website](#) for more about knowledge transfer opportunities.

6.2　Disclaimers and Acknowledgements in Publications

a) Statistics Canada and the Canadian Research Data Centre Network (CRDCN) request that researchers publishing findings from their RDC research include the following notice:

> "This research was supported by funds to the Canadian Research Data Centre Network (CRDCN) from the Social Sciences and Humanities Research Council (SSHRC), the Canadian Institute for Health Research (CIHR), the Canadian Foundation for Innovation (CFI), and Statistics Canada. Although the research and analysis are based on data from Statistics Canada, the opinions expressed do not represent the views of Statistics Canada."

b) In addition, researchers publishing data tabulations *involving Forward Sortation Area* (FSA) geography must include the following statements:

> "©This data includes information copied with permission from Canada Post Corporation."

The following text must be included in the documentation that is provided with any table involving FSA geography:

"The attached data (insert table title) at the FSA level is provided for use in accordance with the terms and conditions of the Statistics Canada Open License Agreement hereby attached. In addition you are required to post the following notice whenever content of (insert table title) at the FSA level is published."

6.3   Copyright

The terms and conditions of copyright are stated in the MRC. Statistics Canada holds copyright on vetted tables released from the RDC. Should Statistics Canada wish to use the tables in a publication, the Principal Investigator will be informed. The Principal Investigator holds copyright on any publications created with the vetted tables.

6.4   Ethics Approval

Many universities have decided that secondary analysis, using Statistics Canada data, does not require an ethics review. However, in some universities an ethics review is necessary. In addition, some journals, especially in the medical fields, require an ethics review in order for the article to be published in the journal. Should a researcher find that an ethics review for his/her RDC research is necessary, the document "Mitigation of Risk to Respondents of Statistics Canada's Surveys"[49] outlines the policies and procedures Statistics Canada has established to mitigate the risk to respondents of Statistics Canada's survey.

---

[49] This document has been reviewed by the RDC Program Manager and Information Management Division at Statistics Canada. It is hoped that the information provided will assist the researcher should an ethics review be required.