Canadian RDC Network

Procedures for Opening and Operating
a Branch RDC

May 2009

(This document should be read in conjunction with
Procedures for Opening a Research Data Centre, May 2009)

TABLE OF CONTENTS

1. Introduction

A special need has been expressed by some universities for an alternative access to Statistics Canada master file data in a secure environment identical to a RDC. These universities do not have the level of funding or the number of researchers necessary to open a full time Centre where at least one Statistics Canada analyst is present full time.

This led to the idea of a "Branch RDC," that is, a secure laboratory controlled by Statistics Canada in which an employee from the agency is always present when the laboratory is open. This employee, a Statistical Assistant, maintains security and works in cooperation with the analyst in the host RDC. This document explains the concept and provides general guidelines on the best practices to manage a Branch.

This document is to be read in conjunction with the document "Procedures for Opening an RDC, rev. May 2009".

Unless there are exceptional circumstances the expectation is that new RDC sites will begin as Branches of an existing Centre and remain as Branches until the level of activity warrants consideration as a full Centre.

For researchers, the Branch RDC can reduce the need for travelling. It also facilitates the cooperation among researchers from the same region or province that are located in different cities or different universities within the city. Compared to a regular RDC, a Branch has fewer workstations, is open part-time (between 8 and 25 hours a week) and typically provides only the most commonly used statistical software. Branches are staffed with part-time statistical assistants. There is no Statistics Canada RDC analyst on site, which means that research output and computer programs are not vetted locally. These outputs are encrypted and securely transferred to the host RDC associated with the Branch. In the host RDC, the Statistics Canada analyst will vet the output before releasing the information to the researchers. The analyst in the host RDC coordinates the administration of the research contracts in the Branch (in collaboration with the statistical assistant) and also provides remote assistance on questions related to survey content. The logistics of this vetting procedure will undergo changes as the Network moves to complete its intranet connections in 2009.

2. Implementation and Operation of a Branch RDC

2-1. Physical Features of Branch RDCs

The same characteristics of physical security are present in the Branch laboratory. Here are the most important ones:
- Entries and exits are managed and monitored by a specialized software.
- Movement detectors are installed.
- An alarm system is linked to university security.
- Researchers have to use an access card.
- When the lab is closed, a key is necessary to enter (in addition to the access card).

- Security of the computer network is maintained by the use of individual user accounts for each researcher/project which are restricted to certain directories.
- Researchers' computers have no active USB ports, diskette or CD-ROM readers.
- An encryption software is installed on the stc employee computer (secure network) for the transmission of unvetted outputs.

The RDC program outlines the complete set of requirements in the document "Security Requirements – 2007 version".

The required equipment for a branch includes a shredder, lockers and a secure filing cabinet (or a safe). The physical protection of the monitors needs to be maintained (no direct view from one workstation to another).

Commonly used statistical software is installed on the researchers' workstations (or available through the server). Data are only stored on the server, not the workstations.

It is possible for researchers to print on green paper but this material must never leave the laboratory.

The university provides IT assistance for the management of the computer network, including regular back-up of the content of researcher folders. The persons involved have to obtain the security clearance and are sworn in under the Statistics Act.

Before a centre can be opened, there is an inspection performed by Statistics Canada. It is conducted by the Chief of Facilities Management, from the Administrative Support Services Division. A senior representative of the RDC program is also present for this inspection.

In a Branch RDC, there is a limitation on the list of survey data offered. Contrary to the practice used generally in the host RDCs for core surveys, a Branch will normally only receive and install the survey data used by their active projects as opposed to the full collection that is placed in full-service RDCs. The transfer of data to a Branch occurs in the same way as a full-service RDC.

Following the guidelines from the Public Works and Government Services Canada, it is expected that the **office space** allocated for each Statistics Canada employee in the Branch RDC is a minimum of **90 square feet**.

2.2 *Management of Branch RDCs*

Branches will create a local management committee composed of representatives from various faculties. This committee can play a leading role in conducting promotional activities and interact with the Management Committee of its host Centre.

A Statistics Canada employee must be present in the Branch when it is open to researchers. Branches are staffed by a **statistical assistant** as opposed to an RDC analyst (in a full-service RDC). The service provided by a statistical assistant differs from that provided by an RDC analyst. For example, no disclosure avoidance analysis is conducted in a Branch. Results are

encrypted and sent by courier to the host RDC for processing by an analyst. The same applies to syntax (as SAS, SPSS or STATA programs). This procedure (the transfer of material between the Branch and the host RDC) will change once the wide area network (WAN) has been implemented in CRDCN.

The orientation sessions are conducted by the analyst in the host RDC. Different options are available: the analyst can invite a new researcher to come to the host RDC, where a session can be organized with other new users there (for an efficient use of the analyst's time). If no orientation session is planned in the host RDC and there is a team of researchers starting in the Branch, the analyst may decide to go to the Branch. In that case the travel cost is covered either by the "Branch" or the host university, depending on the nature of the agreement into which they entered (see Section 3). Other options are described in Appendix 4.

Support for data concepts and content and for methods is provided by the RDC analyst in the host centre. Full documentation of the questions and answers is maintained so that they are available for future reference. It is recommended that all such communication between the statistical assistant and the RDC analyst be in writing (through e-mail).

2.3 Responsibilities of Researchers

Researchers using a branch are subject to the same requirements as those using a Centre. (see document on "Procedures for Opening an RDC, rev. May 2009)

2.4 Recruitment of the Statistical Assistants

The statistical assistant needs to have a completed bachelor degree in demography, economics, sociology or a related field, and an acceptable specialization in economics, sociology or statistics. The ideal candidate must also have some experience in quantitative research. Where possible, she/he can also provide basic assistance to students that have questions on the use of statistical software.

The presence of a Branch has an impact on the workload in the host RDC. The analyst in the main center is providing a remote support to the research team located in the branch. This means that communication has to be done over the phone and by email. Experience has shown that it may necessitate more time to respond to complex requests when the analyst cannot sit beside the researchers. This is especially true for the disclosure avoidance analysis conducted at distance.

Statistical assistants are hired using the standard procedures as defined by the Public Service Commission. The academic director of the Branch is generally invited to participate in the hiring process as a member of the hiring committee. The number of assistants that need to be hired depends on the number of hours that the Branch will operate, as determined by the academic director. Statistical assistants have typically been hired as part-time workers, which limits the number of hours they may work in a week to a maximum of 12.5. The practice in existing labs has been to open between 7.5 and 25 hours a week. Up to now, the staff that have been hired have been mostly graduate students. Frequent turn-over has occurred in the existing Branches (students are graduating and move to full-time positions). When possible,

alternative options can be considered to reduce staff turnover since this is very costly for the Branch, the host university and Statistics Canada.

A statement of merit and a poster have to be used to staff the Branches. The job description is defined in consultation with the academic director. An example is presented at the section 6. The position must be advertised properly on the University campus and on the local RDC web site. Interview material must be prepared in advance and linked to the statement of merit criteria.

Applications are screened and candidates meeting the requirements are contacted for an interview. Generally, the hiring committee is composed of two Statistics Canada employees, the analyst and the supervisor, and a representative of the University, generally the academic director. Reference checks and security check must be completed before an appointment can be made.

The successful candidate is "sworn in" by the RDC analyst and goes through a process of "documentation" with the STC human resources.

The statistical assistant also needs an encryption key in order to communicate with the RDC network. The encryption key and program are provided by Statistics Canada.

The interview travel cost for the RDC analyst and supervisor are covered by Statistics Canada.

2.5 Training and Contact with the Statistical Assistants

As the statistical assistants are generally starting their first position with Statistics Canada, it is essential to provide them with extensive training on all the operational aspects of a secure laboratory and also on the protection of confidentiality. These new employees must visit and become familiar with both the host RDC and the Branch before they start to work.

The RDC Program has developed training material and a checklist of elements that can be used for new staff as the statistical assistants.

The initial phase of training can include two days where the RDC analysts spend time with the assistant. This can be done at the main RDC and at the branch RDC (which has the advantage of developing an awareness of the two environments).

In existing branches staffed with two assistants and where one of them is replaced, the other assistant can also assist in the training conducted by the analyst.

We have to note that the initial training is not complete and has to be followed by a form of continuous training, as specific challenges occur. This can be managed through conference calls, or video-conference if available.

For the training activities, the travel costs of the assistants, analyst and supervisor are covered by the University partners.

2.6 Transfer of Data and Researcher Folders

When the laboratory has been approved according to the security requirements and the statistical assistants have been hired, the transfer of the master file data can begin. This step will be done following an exhaustive testing to verify that the account and folder structures conform to the proper security policies.

The survey data needed for the research contracts which is being moved to the branch needs to be encrypted. The initial transfer is done by an analyst travelling to the Branch and guiding the statistical assistants in the setup of the file structure. The travel cost is covered by the university partners (as agreed to in the signed memorandum of understanding (see Section 3).

Future data updates are processed as follows:
- Head Office encrypts the data on one DVD with all the Entrust keys associated to both the host RDC and the Branch(es).
- That DVD is sent to the host RDC by courier.
- The RDC analyst in the host RDC will install the data in his/her Centre and verify if the data are needed for the Branch.
- If needed, the RDC analyst will transfer the encrypted data to the Branch by courier and email the statistical assistant to notify them of the transmission of data.
- RDC assistant confirms the receipt of the data.
- RDC assistant decrypts the data and moves them to a properly defined folder (this is done in close cooperation with the host RDC analyst and the local LAN administrator).
- These procedures will be modified once the WAN is implemented across the CRDCN.

2.7 Transfer of Results and Syntax

When researchers want to take out **results**, they have to fill-in a disclosure request form and inform the statistical assistant. The assistant will encrypt the results and send them to the analyst by courier in a double-sealed envelope. The vetting will be done in the host RDC. Generally, this only adds one day of processing due to the shipping time. In the host RDC, results are vetted as regular requests. The vetted files can be sent to the researchers by email assuming no issues of disclosure are found. If a problem occurs in the vetting process, the analyst will contact the researcher and keep the statistical assistant informed of the process.

For **syntax**, the researchers inform the statistical assistant and then the files are encrypted and sent to the host RDC (also by courier in a double-sealed envelope). After the content has been verified, the analyst sends the files to the researchers by email.

The assistant provides a general support to the branch, but the responsibility of conducting disclosure control is only performed by the analyst, who has the necessary skills and experience. The statistical assistant reports to the RDC analyst in the host centre. A list of responsibilities for the statistical assistants is included in the Appendix 2.

The courier cost for shipping encrypted files between the Branch and host RDC are covered either by the Branch University or the host (depending on the content of the MOU).

A detailed description on the processing of disclosure requests is presented in appendix 3.

2.8 Parallel use of the Branch and the host RDC

Some researchers using a Branch RDC may consider that their work is limited because the lab is only open part time. In cases where the Branch RDC is not too far from the host center, researchers may ask to use both labs in order to have "five days per week" access.

This cannot be solved by frequent transfers of data between the Branch and the host RDC. It would have negative consequences for the security of data and the workload of the RDC employees.

An alternative solution is that only the syntax travels between the two labs. This means that each center has the source data file and that researchers are modifying these files by running the same syntax (to create new variables and custom files).

From the host RDC to the Branch, the analyst can easily screen the syntax to make sure that no results and no identifiers are present. After that, the syntax file can be sent to the Branch by email (as it is established that the content is not confidential).

If a syntax transfer is needed to move from the Branch to the host RDC, the statistical assistant cannot verify the content. That employee will encrypt the syntax file and copy it on a CD-ROM or USB key. A double-sealed envelope will be used and the researcher will transport it to the host RDC. The assistant in the Branch will notify the RDC analyst that the file is being transferred that way.

2.9 Promotional Role of the Statistical Assistant

If some of their time is available, the statistical assistants may play a role in the promotion of the Branch RDC. This could include meeting groups of graduate students potentially interested in using Statistics Canada data.

If a local Web page is created for the Branch, the assistant can assist the webmaster with the updating of that page.

Statistical assistants can also identify the research groups that could be interested in the content of the detailed social surveys hosted in the RDCs. This work can be done in cooperation with the local management committee.

2.10 Potential Development of the Branch

If the level of activity in the Branch is higher than expected and access to a workstation becomes an issue, then it may be necessary to consider an increase of the opening hours. A Statistics Canada part-time employee can work 12.5 hours a week, so a Branch could be open 25 hours with two employees.

Beyond that level, if a Branch needs to open full time, it would be expected that the centre become a full RDC and staff it with an analyst. This will happen when the volume of contracts, related disclosure avoidance control and the maintenance of a large set of surveys expands.

3. Cost Sharing for the Functioning of a Branch

The host university and the Branch University must enter into an agreement that includes the distribution of costs and tasks related to the operation of the Branch. All agreements concerning the services provided by Statistics Canada are between the agency and the host university. (There is no separate agreement between Statistics Canada and the Branch University)

The following costs need to be covered in the relationship between the Branch and the host RDC. The decision on which partner is responsible for a particular item is to be negotiated between the participating universities. However, the usual practice based on experience to date is shown in parentheses following each item.

- Services provided by the host RDC analyst to the Branch (normally covered by the Branch, but the size of this partnership fee can vary depending upon the level of activity in the Branch).
- Services provided by the Branch statistical assistant (usually paid by the Branch to the host RDC).
- Travel of the Branch statistical assistant to the host RDC for training (usually covered by the Branch).
- Travel of the host RDC analyst to the Branch for the installation of data (usually covered by the Branch).
- Courier service for letters and encrypted material sent between the Branch and the host RDC (usually paid by the Branch).
- An arrangement should be worked out so that the percentage of the allocation formula that the host RDC receives from the Network for its activity and output is allocated to the Branch in a fair proportion.
- For contracts executed in the Branch, the proportion of the contract given to the Branch could equal the percentage of the partnership fee.
- Collaboration between the host RDC and the Branch could include some training opportunities from the host RDC, in the form of workshops for academic researchers and graduate students, fellowship opportunities for graduate students, regional conferences, joint promotional activities, etc.
- A model agreement between an Centre and a Branch is available upon request.

4. Conclusion

The core content provided in this document should stay valid for a certain period of time. It will be updated when necessary. In addition, other elements of information that can be useful for the management of branches are grouped in a series of four appendices. These elements are more subject to change regularly.

Appendix 1 presents a summary of the history of Branch RDCs. Appendix 2 is providing a sample of job offer for statistical assistants. The third appendix describes the process of disclosure control to release output from branches. Finally, the fourth appendix is announcing future developments that are planned.

APPENDIX 1 - History of Branch RDCs

The first four Branches opened with the host RDC being the CIQSS at the University of Montreal. **Sherbrooke University's** Branch opened in October 2004. This lab includes 5 workstations and one office for the Statistics Canada employee. Currently, one statistical assistant is working in this Branch and the lab is open 12.5 hours a week.

The Branch in Quebec City at **Laval University** opened in February 2005. The laboratory has 7 workstations for researchers and an office for the employee. The Branch is open 25 hours a week, with two part time statistical assistants.

Two Branches were opened within the same city as the host RDC at the University of Montreal. The **UQAM Branch** RDC opened in February 2006. This lab has 6 workstations and has been staffed with two part-time statistical assistants. The lab is open 25 hours a week with two assistants.

McGill University's Branch opened in May 2006, and offers 8 workstations. One statistical assistant works 12.5 hours per week. In addition, an RDC analyst who serves the whole CIQSS network is present part-time. The lab is open 25 hours a week.

The Branch RDC at the **University of Victoria** opened in August 2006. It has 6 workstations and is open 16 hours a week, with two part time assistants. The host RDC is the BCIRDC at the University of British Columbia.

Laurentian University in Sudbury opened its Branch in December 2006 as a Branch of SWORDC located at the University of Waterloo.. The laboratory has 2 workstations for researchers and is open 8 hours a week.

Three other Branches will open in 2008. **Simon Fraser** will be a Branch of BCIRDC. **The University of Windsor** will be a Branch of the University of Western Ontario, and **York University** will be hosted by **The Toronto Region Statistics Canada RDC.**

APPENDIX 2 - Sample of Job Offer for a Statistical Assistant

JOB OFFER – STATISTICS CANADA

Social and Demographic Statistics Branch - Research Data Centres Program

Position Title: Statistical Assistant
Position number: …
Group and level:
Tenure: **Part-time worker (12.5 hours per week)**
          Starting date: …
          Duration: 6 months (renewable)
          Salary: This is a part time job. As a reference, the annual full time salary range is from $40,101 to $43,286. The salary is prorated at 25 hours per period of pay of two weeks.
Location: …

Challenge
- Represents Statistics Canada presence in the RDC.
- Opens the RDC (or Branch RDC) to facilitate researchers' access to Statistics Canada data and maintains the integrity of the security systems in the Centre.
- Conducts security clearances for researchers, prepares research contracts and amendments, and coordinates communications with users under the supervision of the RDC Analyst.
- Keeps log of research activities.
- Loads the staged data sets on the server.
- Provides basic assistance to researchers on the use of the computer network and statistical software (SPSS, SAS or STATA).
- Maintains a liaison with the Analysts working in the host RDC, including transmission of encrypted results and follow-up on contract administration.
- Participates in the promotional activities, under the guidance of the academic partners
- Provides information to the webmaster for the local RDC web site.

Education
- Completed bachelor degree from a recognized university in demography, economics, sociology or a related field AND an acceptable specialization in economics, sociology or statistics. A completed master degree is an asset.

Experience
- Experience with computer networks.
- Experience with security system or protocol.
- Previous administrative experience.
- Experience in the use of advanced statistical techniques in SAS, SPSS or STATA, in the conduct of a social science research project.

Knowledge
- Knowledge of Microsoft Office and other software packages.

<u>Abilities</u>

- Ability to manage the daily operations of a Research Data Centre while maintaining the Statistics Canada security standards.
- Ability to communicate verbally and in writing.

<u>Personal suitability</u>
- Effective interpersonal skills, dependability, initiative, judgement.

<u>Reliability/Security</u>
- Enhanced reliability check.

<u>Language requirement</u>
- …

To Apply:
Send your CV and cover letter via email to ... (name, title, STC, email and phone number)

Deadline: …

APPENDIX 3 - Processing of Disclosure Requests in Branch RDCs

1. **Before sending the request, check the following with the researcher:**
   - Has the Disclosure Request Form been completed properly (identification of researcher, name of project, names of files to be released)?

   - Is the researcher following the rules on the minimum number of cases per cell?
     - In general, a minimum of five cases per cell
     - For postcensal surveys, the Workplace and Employee Survey (employer file) and the Longitudinal Survey of Immigrants to Canada, a minimum of 10 cases per cell
     - Check for survey specific confidentiality rules

   - Have the variables been properly identified?

   - Do the results apply to a subsample? If so, has the researcher specified it?

   - Has the researcher attached the files required for analysis?

   - Is the researcher providing weighted and unweighted versions of the results?
     - If the researcher wants to use unweighted results, has he/she provided written justification (justification is required in all cases except when the researcher is using STATA commands that begin with "xt"; such commands do not permit the use of weights)?

2. **Storing the files**
   - 2-1. The researcher is required to save his/her files and the Request Form in the research project's "Results" directory.

   - 2-2. Since files in that type of directory cannot be changed, it is recommended that the researcher prepare the Request Form and files in his/her personal directory and then copy them into the "Results" directory.

   - 2-3. The assistant then files the requests based on the following criteria:

3. **Request_*NO_User_DATE***
   - 3-1. Request number (NO)

   - 3-2. User's name (USER)

   - 3-3. Date (DATE) (year_month_day)

         Example: "Request_01_Fleury_2006_01_27"

3-4. The directory will be structured as follows:

Request_01_Fleury_2006_01_27	
Request.doc	*Completed Request Form*
- Files to be vetted	*Directory containing the files to be removed*
- Consultation files	*Directory containing the files that will not be removed*
- Vetted files	*Directory containing the files sent by the analyst (see section 4)*

4. **Sending the request**
Step One: Zip and encrypt the data

1. Copy the "Results" directory to the desktop

2. Zip the "Results" directory on the desktop

   ✓ Right-click on the directory and select WINZIP / Add to *directory name*.zip

   ✓ With the test option, verify that the compression was successful

3. Encrypt the zip file

4. Right-click on the zip file and choose Secure for List / Select Recipients

5. ENTRUST will ask you for your profile name and password

   Profile name: RDCXXX
   Password: XXXXXX

6. Select recipients and click Add. The selection is as follows:

   i. University of Montreal
   ii. University of Montreal2
   iii. University of Montreal3

7. Click OK.

   With the option Entrust Advanced Test, check that the encryption was successful

Step Two: Burn a CD-ROM or copy to USB key

Check that each file copied on the CD-ROM (or USB key) has a ".ent" second extension in their name (as "test1.zip.ent")

Branches may use this space for their own local procedures for burning CDs.

Step Three: Ship the CD or USB key

Branches may use this space for their own local procedures for preparing packages and shipping them by Purolator.

For a Branch located in the same city as the host RDC, the assistant has another option: give a USB key to the researcher and ask him/her to take it to the host RDC in person. In that case, the procedure is as follows:

1. The assistant places the USB key **in a double-sealed envelope** bearing the name of the responsible analyst at the host RDC.
2. The assistant immediately sends the analyst an e-mail informing him/her of the researcher's name and the fact that he/she will be delivering a disclosure request.
3. The researcher takes the envelope to the analyst.
4. The analyst notifies the assistant by e-mail that the envelope has been received.
5. If the assistant does not receive confirmation of delivery, he/she must check with the researcher to ensure that the disclosure request was delivered to the host RDC.

Step Four: If a request is sent by mail, notify the analyst by e-mail that a request is on its way

1. Give the researcher's name.
2. If there is more than one request for the same analyst, number the requests in order of receipt.

5. **Receiving the request**
    When the analyst has finished processing the request and e-mails the files to the researcher, the Branch that originated the request should be CC'ed.

    1. The assistant takes the files attached to the e-mail and copies them to the disclosure request's "Vetted files" subdirectory.

    2. The researcher may consult those files in the RDC, but under no circumstances is the assistant to forward them to the researcher if the latter so requests. For example, if a researcher loses files that were sent previously and would like to have them re-sent, the request should be forwarded to the analyst.

APPENDIX 4 - Future Developments

To facilitate the remote support to the Branch RDC, some forms of technical assistance can be implemented. For example, the **Video-conference** systems can be used in order to conduct the orientation session. The same system can be used to facilitate the process of researchers having to take the oath of office in the presence of a Statistics Canada indeterminate staff. The RDC analyst could conducts the orientation session from the host Centre with the researcher viewing the presentation in the Branch with the statistical assistant. The researcher can sign the oath at the same time. This procedure saves time and travel costs. It is already in use for Branches in Quebec City and Sherbrooke.

Alternatives to the regular RDCs and the Branch RDCs can be explored too. For example, a part time RDC can be opened which is staffed with an analyst. This may be an option that provides more continuity in terms of human resources. This is especially true as the average duration of the statistical assistants tenure is under 12 months. So an important challenge is to deal with the turnover in the Branch staff.

Other than having an analyst, a more permanent STC administrative assistant position can be considered for the Branches, as for example a regular position with an employee working 3 days a week.

Finally, the network of secure data lines should be completed by 2009. With this **Wide Area Network** (WAN), communications between the Branches and the host Centres will be enhanced.