SECURITY IN THE RESEARCH DATA CENTRES (RDCs)

Statistics Canada data are a rich source of information on many facets of Canadian society. They present an unparalleled base of information on which to conduct policy relevant analyses that can contribute to the betterment of our society. 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 the Statistics Canada offices. The success of this endeavour is in the best interests of all concerned; Statistics Canada, the institutions in which the centres are located, the social science research community, the Social Science and Humanities Research Council, other Federal government agencies and Canadian society.

Security of the data is one of the crucial elements of this project. To quote Dr. Fellegi:

> "At Statistics Canada, the protection of respondent information has always been very much a part of the traditions and culture of the staff." (Memo to all senior managers announcing the release of the Security Practices Manual, 1991)

Statistics Canada takes great care to respect the trust that all respondents place in the organisation to safeguard the privacy and confidentiality of the information that they provide to the agency. It is this trust that makes it possible for Statistics Canada to continue to collect accurate and meaningful information.

The creation of the RDCs represents the first instance in which data that are labelled confidential are removed from the physical premises of the Statistics Canada offices. This project is under the scrutiny of the major stakeholders in the National Statistical System - the respondents. It is also of interest to the federal and provincial privacy commissioners and it has attracted media attention. In effect, the security measures that are implemented in the RDCs must be visible and they must be seen to safeguard the confidentiality of the data to the same degree as in the Statistics Canada offices. It is in this spirit that the following security requirements are presented. The specifications presented below are derived from Statistics Canada's Security Practices Manual and its supporting documentation.

Security in the RDCs is addressed in four ways. First, the "culture" of security referred to by Dr. Fellegi will be developed within the RDCs. Second, all analytical products, whether in electronic form or on paper, will be screened to ensure that disclosure of confidential information does not occur. Third, suitable measures will be implemented to ensure the security of the physical environment. Fourth, suitable measures will be implemented to ensure the confidentiality of the data in the RDCs and to ensure that no unauthorised access is possible. It should be noted that each RDC will be subject to a security audit before any confidential data are transferred to the centre.

Statistics Canada and the researchers employing the confidential data should both take all possible precautions, and be seen to be taking all possible precautions, to prevent the release of confidential data. Respondents providing such data place their trust in the

statistical system, and it is in everyone's interest to be seen to be keeping this trust. The following sections address each of these aspects of security.

1. **Creating a culture of security**

   All employees of Statistics Canada are required to undergo security screening and to take an oath or affirmation of office as part of the hiring process. As part of this process the employee agrees that he or she "… will not without due authority in that behalf disclose or make known any matter or thing that comes to [his or her] knowledge by reason of [his or her] employment as described herein." (Oath of office). The creation of a culture of security begins with the administration of this affirmation or oath and continues throughout the individual's career at Statistics Canada. Courses are provided. Documentation is made available on a regular basis. The Security Practices Manual is available to all employees on Statistics Canada's intranet site. In other words, the importance of protecting the confidentiality of the information that the respondents provide to Statistics Canada's surveys and data collection activities is continually reinforced.

   In many ways the approach that is proposed for the researchers who will have access to the RDCs is similar to that described in the preceding paragraph. All researchers whose projects are approved under the terms of the RDC must be "sworn in" under the Statistics Act in order to obtain access to the data. This amounts to taking the oath or affirmation of office described above. A security check will also be performed on all members of the approved research teams. Finally, all researchers accessing the RDC for the first time will be required to attend an information session that covers the operation of the centre, security measures that will be implemented, the importance of and methods for safeguarding the confidentiality of the respondents' data and general administrative information that will help them to accomplish their goals. A handbook containing this information will be provided to everyone concerned with the project following the information session. The researchers will be required to acknowledge receipt of this information by signing an authorisation form.

2. **Disclosure analysis and confidentiality vetting**

   All data and results to be physically removed from the RDCs are carefully screened to ensure that confidentiality is not violated. Data confidentiality in this context refers to the disclosure of information that can be attributed to individual respondents (e.g., persons, households, businesses, other organizations). 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).

The following examples illustrate the various forms of disclosure.

- 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 Micro-data 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 organisation’s reputation with respect to confidentiality. Damage can 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.

The risks of disclosure vary by the type of result that is produced. For example, tabular output and working data files present a very high risk of disclosure and require very close scrutiny. Conversely, model output tends to present a substantially lower risk of disclosure, although some conditions (such as saturated models or the use of many dummy variables) may increase the risk. It is usually the case that the researchers work together with the RDC analysts to ensure that none of the results to be removed from the RDC present a risk of disclosure.

3. Physical security

Access to the RDCs must be restricted to people who are authorised to be on the premises. This includes the Statistics Canada staff person and all researchers whose projects have been approved by the adjudicating committees. The security specifications outlined below represent the minimum acceptable level for an installation in which confidential data collected by Statistics Canada are to be stored:

- Access to the centres must be limited to the people authorised to be present. This includes the research teams and the Statistics Canada personnel who provide support.
- Adequate safeguards must be implemented to prevent unauthorised removal of equipment or data.
- A continuous record of all persons entering or leaving the Centre must be maintained.
- The construction of the Centres must conform to the following specifications:

    - ✓ Perimeter walls should be either cement, concrete block, or drywall with expanded wire mesh from slab to slab, or a suitable equivalent.
    - ✓ Doors should be solid core or steel with steel frames and tamperproof hardware (e.g. non-removable hinges), or a suitable equivalent.
    - ✓ Outside windows at grade level, or lower than 3m should be secured with heavy duty grill work, or a suitable equivalent.

Access to the centres must be controlled with an electronic card access system. The card access system consists of an ID card that contains electronic information enabling the unlocking of doors and allowing the system to register all access into the area. It may also be possible to install readers on the workstations thereby using the same access method for the physical facility and the data facility. In addition, it is mandatory that a suitable deadbolt lock be installed to complement the card access system.

It is also recommended that a continuing record be kept of all entries to and exits from the RDCs. This function may be automated using the access control system. Many electronic card access systems include a function that also records all entries and exits.

Monitoring during normal operating hours will be accomplished through the presence of a Statistics Canada employee since access by the research teams will be limited to the agreed upon hours of operation. However, it is essential that the RDCs also be monitored outside of the normal hours of operation. This may be accomplished through regular off-hour inspections by authorised security personnel, as it is done in Statistics Canada offices, or through the installation of an intrusion alarm system (motion detectors, window breakage, door contacts, etc. depending on the location).

4. Computer System Security

Access to the workstations and to the data sets will be restricted to authorised users. Workstation and data set access controls will be achieved either through the judicious use of passwords and directory access controls that exist on platforms such as Windows NT, through the use of swipe card readers that are programmed to read the same cards that the authorised researchers to access the physical facility (as described earlier in this document) or through some combination of both approaches.

The general configurations for the networks in the RDCs will respect the following requirements:

- All workstations that are used by the researchers have no local storage capacity (i.e. the hard drive, if present, will be disabled).
- All data will reside only on the server and they will be encrypted and password protected.
- All workstations, other than those assigned to the Statistics Canada RDC analyst, will have no output devices that allow the creation of diskettes, compact disks, digital tapes or other comparable media.
- There will be no connection between the network in the RDC and any other network or service outside the physical walls of the centre.
- There will be no local printers attached to the workstations. All printing will be done on the network printer under the control of the RDC analyst.
- Researchers working in the RDCs will not be allowed to bring any personal computing device (e.g. laptop computers, PDAs, etc.) into the centre.

One final point concerning the use of confidential output in the RDCs needs to be made. In the course of conducting analysis it may occur that the researcher will be required to examine reports and output containing confidential information. It is recommended that provisions be made within the RDCs for an area in which such work may be undertaken (e.g. a conference room, a desk or a table). It is imperative that secure storage facilities be provided in the event that this phase of the analysis extends over a number of days. Printouts containing confidential data must be stored in a locked cabinet when not in use. Finally, it is important that appropriate facilities be included in the centres to dispose of such printed material in accordance with the requirements of the security procedures described in the manuals listed above. The disposal requirements can be satisfied through the installation of an approved paper shredder in each RDC.

LA SÉCURITÉ DANS LES CENTRES DE DONNÉES POUR LA RECHERCHE (CDR)

Les données de Statistique Canada constituent une source abondante de renseignements sur plusieurs aspects de la société canadienne. Elles forment une base sans égale d'information sur laquelle il est possible de fonder des analyses de politique qui sont susceptibles de contribuer à améliorer notre société. La création de centres de données de recherche (CDR) dans les grandes universités du Canada offre la possibilité d'étendre l'analyse de ces données au-delà des limites physiques des bureaux de Statistique Canada. La réussite de cette entreprise servira les intérêts de tous les intéressés, c'est-à-dire Statistique Canada, les universités qui seront l'hôte d'un CDR, les chercheurs du domaine des sciences sociales, le Conseil de recherches en sciences humaines, les autres organismes du gouvernement canadien et la société canadienne en général.

La sécurité des données est l'un des éléments critiques de ce projet. Comme l'a fait remarquer M. Fellegi :

> « La protection des renseignements des répondants a toujours été une tradition et une préoccupation du personnel de Statistique Canada. » (Note de service adressée à tous les cadres supérieurs de SC pour annoncer la diffusion du Manuel des pratiques de sécurité, 1991.)

Statistique Canada prend toutes les précautions possibles afin de respecter la confiance que lui accordent les répondants et de sauvegarder le caractère privé et confidentiel des renseignements qui lui sont confiés. Cette confiance est essentielle à Statistique Canada pour pouvoir continuer à recueillir des renseignements exacts et significatifs.

Avec la création des CDR, c'est la première fois que des données marquées « confidentielles » sortent des installations matérielles de Statistique Canada. Ce projet est de très près par les tout premiers intéressés à notre Système statistique national, c'est-à-dire les répondants. Le lancement des CDR intéresse également les commissaires fédéral et provinciaux à la protection de la vie privée et il a attiré l'attention des médias. À vrai dire, les mesures de sécurité mises en application dans les CDR doivent être manifestes et il doit être évident qu'elles protègent la confidentialité des données aussi efficacement que si ces données se trouvaient dans les bureaux de Statistique Canada. C'est dans cet esprit que sont présentées les exigences ci-après en matière de sécurité. Les spécifications qui suivent sont tirées du Manuel des pratiques de sécurité de Statistique Canada et d'autres documents y reliés.

Les questions de sécurité doivent être abordées à trois niveaux. D'abord, la préoccupation à l'égard de la protection des renseignements dont parlait M. Fellegi doit se refléter dans la mise en place d'une culture organisationnelle de sécurité au sein des CDR. Deuxièmement, tous produits analytiques, qu'ils soient en format électronique ou sur claire, seront vérifiés pour assurer qu'aucune données confidentielles seront divulguées. Ensuite, il faut mettre en œuvre des mesures judicieuses afin d'assurer la sécurité de l'environnement matériel. En quatrième lieu, des mesures s'imposent aussi en vue de

garantir la confidentialité des données stockées dans les CDR en veillant à ce que l'accès en soit réservé uniquement aux personnes autorisées et à ce qu'aucun renseignement confidentiel ne soit divulgué par inadvertance. Il faut remarquer que chaque CDR sera soumis à une vérification de sécurité avant que l'on y transfère la moindre donnée confidentielle.

Statistique Canada et les chercheurs qui utilisent les données confidentielles doivent prendre toutes les précautions voulues, et les prendre de façon visible et manifeste, pour empêcher la divulgation de la moindre donnée confidentielle. Les répondants qui fournissent ces données font confiance au système statistique national et il est dans l'intérêt de tous les intéressés que Statistique Canada et ses collaborateurs montrent ouvertement qu'ils sont dignes de cette confiance. Les sections qui suivent portent sur chacun des niveaux mentionnés plus haut.

1. **Mise en place d'une culture organisationnelle de sécurité**

   Tous les employés de Statistique Canada font obligatoirement l'objet d'un triage sécuritaire et, au moment de leur embauchage, sont tenus de prêter un serment professionnel ou de prononcer une affirmation solennelle. Dans le cadre de cet engagement, l'employé convient de ce qui suit : « [...] je ne révélerai ni ne ferai connaître, sans y avoir été dûment autorisé(e), rien de ce qui parviendra à ma connaissance du fait de mon emploi. » (serment professionnel). Le développement d'une culture organisationnelle axée sur la sécurité commence au moment de cette assermentation et se poursuit tout au long de la carrière de chaque employé de Statistique Canada. SC offre à son personnel des cours sur la sécurité et diffuse périodiquement des documents à ce sujet. Le Manuel des pratiques de sécurité est mis à la disposition de tous les membres du personnel au site intranet de Statistique Canada. Autrement dit, l'importance que SC accorde à la protection des renseignements confidentiels reçus des répondants lors des enquêtes et autres activités de collecte fait l'objet d'un renforcement incessant.

   La démarche proposée aux chercheurs qui auront accès aux données des CDR est semblable en bien des façons à celle décrite au paragraphe précédent. Avant de pouvoir accéder aux données, tous les chercheurs dont les projets seront approuvés aux conditions du CDR devront d'abord être « assermentés » aux termes de la *Loi sur la statistique*. Ce geste obligatoire équivaut au serment professionnel ou à l'affirmation solennelle mentionné plus haut. Tous les membres des équipes des projets approuvés feront aussi l'objet d'un contrôle de sécurité. Enfin, tous les chercheurs qui viennent pour la première fois travailler au CDR dans le cadre d'un projet devront assister à une séance d'information où on leur donnera des renseignements concernant le Centre en expliquant les mesures de sécurité qui ont été mises en œuvre, l'importance de protéger la confidentialité des données fournies par les répondants et de l'information d'ordre administrative qui les aidera de rencontrer les buts de leurs projets. Toutes personnes ayant accès au DCR recevront un manuel contenant ces renseignements après la séance. Les chercheurs accuseront réception de cette information en signant une formule d'autorisation.

2. Vérification de la confidentialité

Toutes les données et les résultats qui doivent sortir physiquement des CDR font l'objet d'un examen minutieux destiné à s'assurer qu'il n'y a pas de violation de la confidentialité. Dans ce contexte, la violation de la confidentialité des données s'entend de la divulgation de renseignements qui peuvent être attribués à des répondants particuliers (p. ex. des personnes, des ménages, des entreprises, d'autres organismes). Il y a trois types de divulgation. Il y a **divulgation de l'identité** quand une personne peut être identifiée à partir des données diffusées, ce qui mène à la communication de renseignements au sujet de la personne identifiée. Il y a **divulgation d'attributs** quand des renseignements confidentiels révélés peuvent être attribués à une personne. Il n'est pas nécessaire qu'une personne particulière soit identifiée ou qu'une valeur particulière soit donnée pour qu'il y ait divulgation d'attributs. Par exemple, le fait de publier une étroite fourchette des salaires des personnes dans une profession et une région données peut constituer une divulgation.

Il faut veiller à examiner toutes les données devant être diffusées. Même si un tableau à lui seul ne divulgue pas de renseignements confidentiels, il peut y avoir divulgation lorsqu'on combine des renseignements provenant de plusieurs sources, y compris de sources externes. Lorsqu'on peut combiner des renseignements diffusés de manière à obtenir des données confidentielles, il y a **divulgation par recoupement** (c.-à-d. qu'on peut déduire les données supprimées dans un tableau à partir d'autres tableaux).

Voici des exemples des divers types de divulgation :

- Un personnage bien connu, par exemple un sportif professionnel, est sélectionné pour participer à une enquête et les renseignements publiés sur sa collectivité, par exemple le revenu le plus élevé déclaré dans cette collectivité, ont presque certainement été déclarés par lui. (Divulgation de l'identité).

- Les résultats d'une enquête longitudinale mettent en évidence dont le profil de migration sort de l'ordinaire, ce qui mène à son identification. (Divulgation de l'identité.)

- Les parents d'un adolescent de 16 ans sélectionné dans l'échantillon voient un tableau qui montre que tous les répondants de 16 ans qui ont fourni des données dans leur région ont consommé de la drogue. (Divulgation d’attributs.)

- Un article de journal rapporte les propos d'un veuf de 37 ans qui se plaint d’avoir eu à participer à une enquête; or, les totalisations croisées des données de l'enquête révèlent que seulement deux veufs âgés de 30 à 39 ans ont été échantillonnés. (Menant ultérieurement à une divulgation de l'identité et(ou) d’attributs.)

- En combinant plusieurs résultats, une personne repère des renseignements qui ont été exclus à dessein du fichier de microdonnées à grande diffusion parce qu'ils présentaient un trop grand risque de divulgation (p. ex. le pays de naissance des nouveaux immigrants).

Même l'apparence de divulgation peut porter atteinte à la réputation de respect de la confidentialité des données d'un organisme de statistique. Ce peut être le cas dans les deux premiers exemples même s'il s'avère que la mauvaise personne ou le mauvais ménage a été identifié. Réfuter une identification erronée peut avoir pour effet d'accroître le risque d'identification des véritables répondants.

Les risques de divulgation varient selon le type de résultat produit. Par exemple, les tableaux et les fichiers de travail présentent un risque de divulgation très élevé et doivent être examinés minutieusement. Par contre, les modèles présentent généralement un risque de divulgation nettement plus faible, bien que certaines conditions (comme les modèles saturés ou l'utilisation de nombreuses variables nominales) puissent accroître ce risque. Habituellement, les chercheurs collaborent avec les analystes du CDR pour garantir qu'aucun des résultats qui doivent sortir du CDR ne présente de risque de divulgation.

3. La sécurité physique

L'accès aux CDR est réservé uniquement aux personnes autorisées à se trouver sur les lieux, ce qui comprend le personnel de Statistique Canada et tous les chercheurs dont les projets sont approuvés par les comités de sélection. Les spécifications énoncées ci-après illustrent le niveau minimum acceptable de sécurité pour des installations où doivent être stockées des données confidentielles recueillies par Statistique Canada.

- L'accès aux CDR doit être réservé uniquement aux personnes autorisées à se trouver sur les lieux, ce qui comprend les équipes de recherche et le personnel de Statistique Canada qui apporte son appui aux chercheurs.
- Toutes les mesures adéquates doivent être prises afin d'empêcher que des appareils ou des données ne soient retirés des CDR sans autorisation.
- Chaque CDR doit tenir un registre permanent de toutes les personnes qui y entrent ou qui en sortent.
- La construction des Centres doit se conformer aux spécifications suivantes :
    - ✓ Les murs périmétriques doivent être construits en ciment, en blocs de béton ou en cloison sèche à treillis métallique expansé complet, ou être d'une qualité équivalente convenable.
    - ✓ Les portes doivent être des portes en bois à âme massive ou des portes d'acier à cadre d'acier et à quincaillerie inviolable (par exemple, des charnières inamovibles), ou être d'une qualité équivalente convenable.
    - ✓ Les fenêtres extérieures au niveau ou à moins de trois mètres du sol doivent être protégées par un grillage métallique très résistant ou par un autre moyen équivalent convenable.

L'accès aux CDR doit être contrôlé par un système de cartes électroniques. Ce genre de système est fondé sur une carte d'identité, laquelle carte contient des renseignements électroniques qui permettent au titulaire de déverrouiller les portes et au système d'enregistrer toutes les entrées et les sorties par les portes en

question. Il est possible que l'on puisse aussi installer des lecteurs de cartes d'identité aux postes de travail et utiliser ainsi la même méthode pour l'accès aux installations et aux données. En outre, il faut installer une serrure à clef qui viendra renforcer le contrôle de l'accès par cartes électroniques.

Nous recommandons également que chaque CDR tienne un registre permanent de toutes les entrées et sorties de personnes. Cette fonction peut être automatisée à l'aide du système de contrôle de l'accès. Un bon nombre de systèmes d'accès par cartes électroniques comportent une fonction d'enregistrement de toutes les entrées et sorties.

Pendant les heures normales d'ouverture des CDR, c'est-à-dire les seuls moments où les équipes de recherche pourront y avoir accès, la surveillance sera effectuée par un employé de Statistique Canada qui doit être présent pendant toute la période. Toutefois, il est essentiel aussi d'assurer la surveillance des CDR en dehors des heures normales d'ouverture. À cette fin, on pourra faire appel à des inspections périodiques par le personnel de sécurité autorisé, comme on le fait déjà dans les bureaux de Statistique Canada, ou à un système de détection d'intrusion (détecteurs de mouvement, de bris de fenêtre, de contact avec les portes, etc. selon le lieu surveillé).

4. Sécurité du système informatique

L'accès aux postes de travail et aux ensembles de données doit être réservé uniquement aux utilisateurs autorisés. Cet accès peut s'opérer par un recours judicieux aux mots de passe et aux contrôles d'accès aux répertoires, des mécanismes dont sont déjà munis certaines plates-formes telle que Windows NT; à l'aide de lecteurs de cartes magnétiques qui seraient programmés de façon à lire les mêmes cartes que celles dont se servent les utilisateurs autorisés pour entrer au CDR (tel que décrit plus haut dans le présent document); ou par une combinaison quelconque de ces deux méthodes.

Les configurations des réseaux d'ordinateurs dans les CDS doivent rencontrer les exigences suivantes :

- Les postes de travail auxquelles les chercheurs auront accès ne contiendront aucune capacité de stockage local (c.à.d. le disque dur sera débranché, s'il y à lieu).
- Les données ne seront stockées que sur le serveur du réseau local. De plus, elles seront encodées et protégées par un mot de passe.
- Tous postes de travail, à part ceux réservés aux analystes de Statistique Canada, n'auront aucun moyen d'enregistrer des données ou des résultats sous forme magnétique, tel que les disquettes, les disques compactes, le bandes digitales, etc.
- Le réseau local dans le CDR n'aura aucun lien avec n'importe réseau ou ordinateur à l'extérieur du centre.

- Aucune imprimante ne sera branchée aux postes de travail. L'impression des résultats se fera que sur l'imprimante branchée au réseau, et ceci sous le contrôle de l'analyste de Statistique Canada.
- Les micro-ordinateurs personnels, tel que les ordinateurs portatifs et les APN, sont interdits dans les CDR.

Il convient de souligner un dernier point concernant l'utilisation des données confidentielles de sortie dans les CDR. Il peut arriver que le chercheur, en effectuant une analyse, ait besoin d'examiner des rapports et des données de sortie contenant des renseignements confidentiels. Pour parer à ces éventualités, nous recommandons que les CDR prennent des dispositions afin de réserver un endroit à part (par exemple, une petite salle de conférence, un bureau, une table de travail) où ces tâches pourront être exécutées. Nous recommandons aussi que les CDR prévoient l'aménagement d'installations de stockage sécurisées pour les cas où cette étape de l'analyse durerait plusieurs jours. Dans les moments où elles ne sont pas utilisées, les copies imprimées contenant des données confidentielles doivent être rangées dans un classeur fermé à clé. Enfin, il est important que les CDR disposent des installations nécessaires pour détruire ces documents imprimés, le cas échéant, en conformité avec les exigences en matière de sécurité décrites dans les manuels énumérés précédemment. Ces exigences peuvent être remplies au moyen d'une déchiqueteuse à papier, de modèle approuvé, qui serait installée dans chacun des CDR.