

## DATA STEWARDSHIP AT STATISTICS CANADA

Gordon Brackstone and Pamela White  
 Statistics Canada, Ottawa, Ontario, Canada K1A 0T6

## 1. Introduction

The collection and management of statistical data are central to the mandate and functioning of any statistical agency. The collection of data, by whatever means, intrudes to some extent on the privacy of respondents. The agency that collects data has to consider the burden and intrusion it places on respondents against the value of collecting the information it seeks. Once data are collected, the subsequent control of their access and use must accord with the usage and confidentiality undertakings that were provided to respondents at the time of collection. The policies, procedures and practices governing the collection and use of statistical information are a crucial component of the overall management of a statistical agency. This paper describes the approaches used in this regard by Statistics Canada.

Three principal statutes govern the activities and behaviour of Statistics Canada in this domain. The *Statistics Act* (1985) prescribes the mandate of the Agency, its role in the federal government, its powers and responsibilities, and its operating structure. Central to the Act's provisions is an implicit *social contract* with respondents under which the Agency may burden respondents with requests for information, and in some cases demand response, in order to provide information that is clearly of broad public benefit, but with an absolute undertaking to protect the confidentiality of identifiable individual responses. The Agency must continually balance the public benefit of statistical information against individual rights to privacy.

The *Privacy Act* (1985) and the *Access to Information Act* (1985) are laws of general application that apply not only to the activities of Statistics Canada but to all federal government organizations. The *Privacy Act* protects the privacy of individuals and provides them with the right of access to their personal information that is held by federal organizations. In the context of statistical activities the Act provides rights to individuals to know why personal information is being collected and for what uses. It includes protection against unwarranted information collection intrusion and the right to know of, and have access to, their personal information. Informed consent is an integral component of the Act.

The *Access to Information Act* establishes a right of access to information kept in government records. It also establishes the principle that government information should be made available to the public, with exceptions to this right being limited and specific. For example, confidential information collected under the *Statistics Act* cannot be accessed by means of the *Access to Information Act*; however, access to already produced tabular non-confidential information could be requested under the Act if it were not already available to the public through the regular publishing program or by special request. There is no requirement in the Act for Statistics Canada to create tabulations on request outside the cost-recovery régime currently in place.

In summary, the *Statistics Act* provides the legislated authority for the Agency's statistical research activities. The conduct of these activities is guided by the *Privacy Act*. These two Acts define *confidentiality* and *privacy* as these terms pertain to Statistics Canada's activities.

*Privacy* is properly defined as being the right to be left alone, to be free from interference, from surveillance and from intrusion. In the statistical context it provides respondents with protection against intrusive enquiries and some controls over information about themselves. *Confidentiality* is a property accorded to data about individual units that prevents them being revealed in an identifiable form to any unauthorized person. *Security* refers to arrangements that organizations use to prevent confidential information from being disclosed or improperly accessed or used.

These two Acts place formidable responsibilities on Statistics Canada, particularly with respect to the manner in which it collects, processes, stores, uses and disseminates statistical information. The obligations that fall on the Agency as a result of the *Statistics Act* and the *Privacy Act* can be separated into three major categories:

- the need to minimize the intrusiveness of the Agency's statistical enquiries;
- the need to protect the confidentiality of the data it receives;
- the need to control the use of the data it holds.

This paper will discuss these obligations and describe how the Agency exercises its *stewardship* of data. *Stewardship* is not a term commonly used at Statistics Canada in this context, but it admirably captures the sense of being entrusted with the care of something of value, belonging to someone else, and using it wisely while guarding its value and respecting its owners' rights.

Section 2 summarizes some pertinent provisions of the *Statistics Act* and of the *Privacy Act*, and also draws a contrast between statistical use of records and administrative use of records. This distinction is particularly pertinent to the application of the *Privacy Act* to statistical uses of records. Sections 3 to 5 examine each of the three obligations in turn and describe the policies and procedures used at Statistics Canada to ensure that each is fulfilled. Section 6 reviews the recent institution of privacy impact assessments and their effect on Statistics Canada.

## 2. The Legislative context

### 2.1 The *Statistics Act*

The *Statistics Act* provides Statistics Canada with the authority to collect, compile, analyze, abstract and publish statistical information. It grants the Agency the right to examine an exhaustive list of topics ranging from demographic, social, and cultural phenomena to the activities of business and government. It requires the Agency to take

the census of population and the census of agriculture. It provides the Agency with the authority to access administrative records for statistical purposes with specific mention being made of the Agency's right to obtain customs and tax information.

The Act permits the conduct of surveys on a mandatory basis or on a voluntary basis (except for the censuses of population and agriculture which must be mandatory). It establishes the requirement of confidentiality of identifiable individual information and prescribes penalties for violation of this requirement by employees.

To facilitate the coordination of Canada's statistical system, the Act provides for joint collection to be undertaken with properly constituted provincial statistical agencies and for sharing of the resultant returns. It also provides for joint collection with other organizations, but provides respondents with the opportunity to refuse the sharing of their data with the other organization.

Under the Act, the Chief Statistician is given discretion to release certain items of information that otherwise would be confidential. These include information on non-commercial institutions such as hospitals and libraries, information that a respondent has agreed can be released, information on names, industry and employment size group of businesses, and information already public under other statutes.

The Act also permits certain groups of individuals to have access to confidential records for the purpose of providing a product or service to the Agency. It is in this context that the Agency has been able to encourage the analysis by academic researchers of complex longitudinal micro-data files that are currently being made available for approved projects at university Research Data Centres (see Section 5.2).

## 2.2 The *Privacy Act*

The *Privacy Act* applies only to personal information, defined as information about an identifiable individual, and only to personal information under the control of a federal government institution, which includes Statistics Canada. It provides that such information shall be used only for purposes consistent with those for which it was originally collected or for specific purposes listed in the Act, unless the individual to whom the information relates has consented to an additional use. One of the specific purposes listed in the Act is research or statistical work (sub-section 8.2(j)).

The Act also requires that an index of all personal information holdings be published annually. This publication entitled *InfoSource* lists the files containing personally identifiable information. Under the Act, individuals have the right to access personal information kept about them, to request correction of any errors or omissions and to attach a notation to any record for which a correction has been requested but not made, where the record is to be used for administrative purposes.

Respondents have the right to obtain a copy of information that they provided to Statistics Canada. This right is exercised most frequently in the context of the Census of Population. For example, about 100 respondents have requested a copy of their 2001 Census information.

The Act also creates the position of Privacy Commissioner, as an appointee of Parliament, and provides the incumbent with the authority to investigate complaints and to initiate complaints of his or her own. With respect to Statistics Canada, the Privacy Commissioner has in the past investigated a small number of complaints lodged by respondents who are of the view that Statistics Canada has released their personal information without their permission. One such complaint, while ruled as unfounded by the federal Privacy Commissioner, caused the Agency to investigate in greater detail its dissemination practices as they pertain to groups who may feel that their information has been released in sufficient detail to reveal individual characteristics. This issue is discussed in Section 4.6.

As already noted, the impact of the federal *Privacy Act* on the Agency has been considerable. The Agency's Policy on Informing Survey Respondents and its Policy on Record Linkage both incorporate principles to be found in the *Privacy Act*. The specifics of these policies are discussed in greater depth later in the paper. More recently, the federal government has attempted to modernize its application of the *Privacy Act* by developing a policy on Privacy Impact Assessments which requires that federal departments assess and evaluate the privacy impact of new programs that may involve the collection of personal information.

## 2.3 Administrative and statistical uses of records

The distinction between administrative use and statistical use of records is crucial to the management of statistical data, but may be poorly understood in the population at large. An administrative use of records is one in which the objective is to make a decision about the individual to whom the record pertains. This may be, for example, a decision about eligibility, entitlement or qualifications for participation in a government program such as Old Age Security or Employment Insurance. The identity of the subject and collection of personal information, including a personal-identifier such as a social insurance number, is paramount to the decision-making process involved in the administration of such programs.

The objective of the statistical use of records is to describe and analyze characteristics of a population of individuals in which the particular record falls. For example, the production of a tabulation of the number and characteristics of individuals qualifying for a government program or enrolled in a particular university. In this instance, the identity of the individual is of no relevance, only their pertinent characteristics need to be known.

This is a distinction between uses of the information, not a distinction between records. A given set of records could, in principle, be used for either purpose. In fact, the analysis and statistical use of records that were originally collected for administrative purposes, such as tax data, birth and death registrations and customs declarations, is a well-established and legitimate practice that occurs in many countries. This type of statistical activity is encouraged as a means of avoiding duplicative collection and thereby contributing to a reduction in the overall survey response burden imposed by statistical agencies.

On the other hand, the administrative use of records that were originally collected for statistical purposes is forbidden in most countries. The confidentiality provisions under which statistical data are collected preclude their use for administrative purposes which require, by their nature, identification of individuals and result in decisions being made about those individuals. Under most privacy legislation such a use would be considered inconsistent with the purposes for which the information was collected.

On the surface, the presence of identification information in a record may seem to be a distinguishing feature between a statistical database and an administrative database. But two considerations complicate the situation:

- even though identifiers are not required in deriving statistical data from databases, they are sometimes required in the construction and verification of statistical databases;
- even if identifiers are not included in statistical databases, the identity of individuals can sometimes be deduced from the combination of characteristics retained in their record.

Therefore, eliminating the identifiers from a statistical database as a means of ensuring that the database cannot be misused may be both damaging to statistical use and insufficient for the protection of confidentiality. The presence or absence of identifiers is in itself an insufficient indicator of the uses to which a data set may be put.

The use of administrative records for statistical purposes is pervasive in the Canadian statistical system for reasons of efficiency and reduction of burden. A constant challenge for Statistics Canada is to emphasize the value and legitimacy of such use, while assuring the public that there is not, and cannot ever be, any flow of individual information in the opposite direction - from the statistical agency to an administrative body.

It is important to note however that other government jurisdictions are increasingly requesting that the public be made aware of the research and statistical uses to which their administrative information is being put. For example, the *Alberta Health Information and Privacy Act* requires that a privacy impact assessment be completed before administrative health records can be used for statistical and research purposes.

### 3. Minimizing intrusion

The intrusiveness of our statistical enquiries is a subjective attribute determined by our respondents. Four principal characteristics of our enquiries may constitute prime determinants of intrusiveness.

The volume of questions asked, the joint effect of the number of questions on our questionnaires and the sample size of our surveys, may be thought of as a measure of the overall burden imposed by our enquiries on respondents. The pure volume of this burden may be a source of intrusion for respondents.

The sensitivity of questions asked may be a more focused determinant of intrusiveness. A single question on a topic a respondent does not want to reveal may be considered more intrusive than a whole questionnaire of innocuous questions.

The manner in which questions are asked may be an important factor in how intrusive they are seen to be. The mode of response - whether there is a face-to-face interview, a

telephone conversation, a self-completed paper questionnaire, or an electronic interface - can be a prime determinant. In the case of interview surveys, the characteristics of the interviewer may be an important factor (and in the extreme case of an interviewer known to the respondent, an overwhelming factor). The behaviour and competence of the interviewer, the convenience of the mode of response, the clarity of the questionnaire, and the information provided with the questionnaire on the authority, purpose and use of the survey, will also be relevant factors in the respondent's assessment of intrusiveness.

Finally, the discretion the respondent has to refuse to answer will also serve to determine intrusiveness. Mandatory surveys may be seen as more intrusive. Voluntary surveys where the respondent can make an informed decision whether to participate may, other things being equal, seem less intrusive.

In taking account of all these factors when trying to minimize intrusiveness, the alternatives to the enquiries have to be considered. If the alternative is not to do the survey, the alternative is clearly less intrusive. If the alternative is to use information previously collected for a different purpose, assuming that is permissible, the intrusiveness of this alternative has to be considered. The *Statistics Act* and *Privacy Act* specify what is permissible in this regard. A judgement has to be made on whether a permissible alternative is less intrusive than a new direct enquiry. If the alternative is the linkage of existing records from different sources an even more careful assessment of intrusiveness and the balance of benefits has to be considered.

In the following sections we examine some of the means used at Statistics Canada to control intrusiveness.

#### 3.1 Minimizing what we ask

One of the principal ways in which Statistics Canada avoids surveys and survey questions is through the use of information that has already been collected for administrative purposes. The prime example is tax data collected by the Canadian Customs and Revenue Agency (CCRA), but other long-standing examples include birth and death registrations for vital statistics and customs information for international trade statistics.

Tax information is increasingly used by both the business and household survey areas in Statistics Canada. Sections 24 and 25 of the *Statistics Act* permit the Chief Statistician to have access to income tax, customs and excise records for statistical and research purposes. Since 1995, respondents to the longitudinal Survey of Labour and Income Dynamics (SLID) have been asked for their consent to use by the Agency of their income tax records as an alternative to responding to the detailed income questions on the survey. Approximately 80% of SLID respondents give the Agency permission to use their tax data in this way. Income tax returns from individuals are also used to produce annual demographic and income profiles for small areas. A program to make greater use of records from the Goods and Services Tax (GST) to reduce the burden of both monthly and annual business surveys is underway.

Statistics Canada's business statistics program makes extensive use of tax information in place of collecting detailed business financial data via surveys, especially for small businesses. This endeavour has been generally welcomed by businesses who appreciate the significant reduction in response burden that it entails. This activity has been further advanced by recent changes to the federal *Income Tax Act* which has allowed Statistics Canada to share business tax data with provincial statistical agencies having similar statistical collection powers and confidentiality protections.

Data already collected in surveys may also have further statistical use. For example, the program of post-censal surveys makes use of information collected in the Census of Population to identify frames for surveys of designated sub-populations, such as those reporting Aboriginal identity or disability in the census. In such cases, respondents are informed of this intended use at the time of the census. As a result, basic demographic and socio-economic information collected in the Census does not have to be collected again in the post-censal survey.

Opportunities to combine or integrate surveys may also arise. The questions once asked by the annual household income survey have now been integrated into the longitudinal income survey (SLID). In the business area, many annual surveys of different industry sectors have been integrated into one unified enterprise survey. This integration, coupled with an expanded use of business tax data, has been used to achieve expanded industry coverage without adding to respondent burden. It has also resulted in the elimination of a number of business surveys.

In general, before launching any new survey, a review of existing data sources is required and a justification for questions should exist. Agency management will seek strong justification for the data requirement especially if the survey content includes sensitive topics, or explores topics or areas never before addressed in a Statistics Canada survey. The Canadian Community Health Survey<sup>1</sup> (CCHS) cycle on Mental Health provides a good example of the requirement to fully inform not only senior management but also the federal and provincial privacy commissioners, privacy advocacy groups, health practitioners, patients and members of the interested general public of the critical requirement by medical practitioners and the health sector for such sensitive information. In this case, not only was a Privacy Impact Assessment conducted (see Section 6) but cross-country consultation with user and advocacy groups was undertaken and media sessions were held before the content was approved.

Once the absolute necessity of conducting a survey is established, there is still the issue of how many respondents need to be burdened with the request or requirement to respond. Statistical sampling and estimation techniques aim to find the optimum balance between statistical accuracy requirements and cost. Since cost is primarily a function of sample size, these techniques enable us to minimize sample size for a given level

of accuracy. In cases where, for reasons of cost or respondent burden, the sample size cannot be large enough to achieve desired accuracy requirements (e.g. the case of estimates for small areas or detailed industries), the option of statistical modelling may also be available. These statistical methods are not the subject of this paper. Sarndal et al (1992) provides a thorough review, while Rancourt and Hidiroglou (1998) describes an example from Statistics Canada of the use of administrative data, statistical modelling, and survey sampling in combination to achieve a significant reduction in respondent burden in a business employment survey.

In some cases, linkage of data from different statistical or administrative sources may be a potential source of the required information. However, privacy implications arise from the linkage of records coming from different sources and this topic is addressed in more detail in section 5.1.

### 3.2 Reducing respondent burden through joint data collection activities and data sharing

While Statistics Canada has complete control over the respondent burden it imposes, it has less control over the burden imposed by the statistical collections of other organizations. However, it does have some mechanisms available to help in its role of integration and avoidance of duplication in the Canadian statistical system.

Section 11 of the *Statistics Act* allows for joint data collection with a statistical agency of a province having the statutory authority to collect the information and confidentiality protection requirements equivalent to those in the *Statistics Act*. This section is applied primarily in the context of mandatory business surveys. In these instances the business is required by law to answer the survey and is informed that the information collected will be shared with the provincial statistical agency. Provinces avoid the cost of conducting their own survey programs and respondents are spared the burden of responding twice. This section is not applied in the case of the voluntary household surveys to avoid generating survey refusals from those who do not wish to share their information.

Section 12 of the *Statistics Act* permits a joint collection data sharing agreement with any department, municipality or corporation. However, in this case respondents must give their consent to the sharing of their survey information with the named data sharing partner(s), otherwise their response cannot be shared. This section is used for joint collection of certain surveys with federal government departments, and for agreements with provinces on certain voluntary household surveys. For example, the CCHS (cycle 1.1) achieved a data sharing rate of 95%. As the data collection is carried out under the authority of the *Statistics Act*, all of the provisions of the Act apply. Thus, when the survey responses of those who agreed to sharing are provided to the data sharing partner, the partner must abide by the very same strict rules of confidentiality and must ensure that tight security measures are in place to protect the shared file.

### 3.3 Mandatory or voluntary surveys?

Prior to 1981 surveys conducted under the *Statistics Act* had to be mandatory. In 1981, an amendment was passed

---

<sup>1</sup> The Canadian Community Health Survey is a large cross-sectional household survey aimed at producing estimates of health status and health system usage for all 136 health districts in Canada.

allowing surveys (other than the Censuses of Population, and Agriculture) to be conducted on a voluntary basis. Section 8 of the Act requires that a specific order is required to make a survey voluntary, in which case the provisions dealing with penalties for refusal do not apply.

The Agency developed guidelines (Statistics Canada, 1997) for determining when surveys should be voluntary and when mandatory. The guidelines required programs to consider whether the survey results are required to fulfill Federal or Provincial statutes or regulations, whether they would have a direct influence on transactions of domestic and international financial and commodity markets, whether the survey would provide the only source of information needed by government as input to the maintenance of existing programs, or the development of major new programs or policies, and lack of data would compromise significant program objectives, or whether the survey is a basic component of statistical programs meeting the above criteria - for example making a substantial contribution to the System of National Accounts.

The application of these criteria has led to the situation today where most business surveys are conducted on a mandatory basis due to their key role in supporting the system of national accounts or other key economic indicators, or their use in the calculation of federal-provincial transfer payments. Aside from the Census, the only mandatory survey on the household side is the Labour Force Survey which provides the critical employment and unemployment indicators each month and is used to determine eligibility thresholds for receiving employment insurance benefits in different regions.

Most of the Agency's household surveys depend on the voluntary participation of respondents. In the case of a voluntary survey, no penalty for failure to participate can be exercised by the Agency. Participation in these surveys must depend on respondent goodwill and our success in persuading them that the public benefit arising from the collection of such information justifies their individual sacrifice of time and disclosure of personal information. The promise of complete confidentiality of their records (unless they give permission for such a disclosure to a named data sharing partner), and disclosure protection in statistical outputs, aim to allay respondent fears that personal harm or disadvantage might arise from responding. Of course, these explanations and commitments are equally important for mandatory surveys if the cost of the non-response follow-up effort is to be kept in bounds.

Following some earlier criticism by the Office of the Privacy Commissioner of letters and brochures used by the Agency to inform respondents of the voluntary nature of the survey, the Agency has adopted a more explicit wording that must be used in all communications with respondents on this subject.

### 3.4 Facilitating response

The previous sections addressed ways that a statistical agency can seek to reduce the response burden and the intrusiveness of its survey program. They also identified the importance of convincing respondents that participation in the survey is a justified civic duty. Yet, in an increasingly difficult survey environment, statistical agencies must continually ask

themselves if such measures are sufficient. Not only are respondents more difficult to contact and under greater bombardment by public polling and market surveys, but a greater sensitivity to privacy intrusion is creating further barriers. In this environment, a battery of measures is needed to keep response rates high while respecting the privacy rights of respondents.

The need to ensure that respondents are fully informed of the uses to be made of the information they provide, of the reasons why questions are asked, and of the protection afforded to their responses, as well as similar requirements of the *Privacy Act*, have been formally embodied in the Agency's Policy on Informing Survey Respondents (Statistics Canada, 1998). This policy requires that respondents be informed of "the expected use of the statistics to be produced from the survey, the authority under which the survey is taken, confidentiality protection and any data-sharing agreements." It is equally important that interviewers also be fully aware of this information. In personal or telephone interview surveys they play the key role in eliciting responses and must be fully trained to respond to questions about the reasons for the specific survey. Interviewers also receive general training in interview techniques and dealing with respondents.

Informative, timely and well crafted survey respondent materials also serve to encourage a high response rate. Our experience has indicated that when respondents are fully informed of the reasons for the data collection, especially when the information is likely to have direct impact on their lives or the well being of their community, response rates are encouragingly high. For example, surveys about health, literacy and child development, though often very burdensome and involving up to an hour of a respondent's time, have high response rates. These surveys typically have excellent respondent relations materials. As well, they have the support of community members including major advocacy groups and associations.

Working with associations can result in considerable goodwill on the part of respondents as well as public pronouncements of support for the activities of the Agency. This support can also be useful in justifying content to respondents, in discussions with the Office of the Privacy Commissioner, and in highlighting findings during the dissemination period. Given the privacy sensitive environment, and the absolute requirement to produce information that will address a significant program, policy or research issue, the active support and contribution of associations, advocacy and community groups is of paramount importance.

Another key element of encouraging response is the provision of modes of response that suit the respondent. Traditional methods of mail, telephone, and personal interview options are today being increasingly supplemented with electronic options, including response via the Internet. The provision of response options appropriate to the capacity and preference of the respondent, and to the content of the survey, is an aspect of survey design critical to minimizing the burden of our surveys. Whatever the mode of response, but especially for electronic means of collection, stringent

physical and IT security standards to safeguard the integrity of the information collected and protect its confidentiality must be in place and provide assurance to the respondent that their information will remain confidential and secure.

Finally, the questionnaire itself is a crucial element of the response process. Rigorous testing of each questionnaire is necessary before it is administered in a survey. Statistics Canada's Policy on the Review and Testing of Questionnaires (Statistics Canada, 2002) lays out this requirement. The Questionnaire Design Resource Centre within Statistics Canada supports programs in the implementation of this Policy which has led to significant improvements in survey instruments with resultant benefits to data quality, reductions in response burden and improvements in respondent relations.

Survey questionnaire testing, which includes qualitative as well as quantitative measures, can lead to the identification of potential content problems especially as they relate to the asking of sensitive questions and cultural differences in the understanding of survey questions. These techniques can also be used to evaluate the efficacy and readability of introductory letters, brochures and other respondent relations materials. The Working Group on Group Privacy (see Section 4.6) recently recommended that the Agency undertake qualitative testing of statements made by the Agency on topics such as confidentiality, privacy and security so as to obtain a better understanding of how respondents react to these messages.

#### **4. Protecting data confidentiality and ensuring security**

Section 17 of the *Statistics Act* places a strict obligation on the Agency and its employees to keep all individual information obtained under the Act confidential. Exceptions to this restriction are few and carefully circumscribed. The Agency has in place a number of internal policies regarding the collection, use and disclosure of statistical information. Many of these policies contain implementation measures designed to assist staff to understand their roles and responsibilities. The Policy on the Security of Sensitive Statistical Information (Statistics Canada, 1988) defines the level of security to be provided to such information and the responsibilities for ensuring that this level is achieved. Directors play a very important role with their responsibility for controlling and protecting all sensitive statistical information obtained by their respective divisions in the pursuit of their program objectives. For example, they must ensure that appropriate control measures are in place in their division regarding access to confidential microdata files. They must also determine the need to retain identifiable files and to ensure that such files are referenced as required under the *Privacy Act* and *Access to Information Act*.

The following sections describe many of measures that the Agency takes to ensure that its obligation to ensure the confidentiality of respondent information is fully met. These measures include employee training, physical security, electronic (IT) security, disclosure protection in the Agency's statistical products, and stringent processes for managing the exceptions to the confidentiality provision. The final section discusses the concern that collective harm might be caused to a group of people by release of non-confidential information about a group.

#### **4.1 Employee obligations and training**

The unlawful disclosure of confidential information by employees of Statistics Canada is punishable by fine and/or prison sentence. Any employee who uses confidential information for the purpose of speculating in any stocks, bonds or other security is also liable for an offence on summary conviction that carries penalties of fine and/or imprisonment.

The penalties are severe and the Agency endeavours to inform all new employees of their obligations when they swear the Oath under section 6 of the Act. In response to the influx of new employees hired over the past few years, the Agency has taken measures to improve its training in the areas of privacy, confidentiality and security. The flagship training courses that are taken by all new professional employees, such as the Survey Skills Development Course and Business and Enterprise Survey Training contain a module on these three items. The orientation course that all new employees can attend also includes basic training in confidentiality and security.

More recently, a two day course designed for middle managers has focussed on the areas of the application of the *Statistics Act*, *Privacy Act* and *Access to Information Act*, disclosure avoidance, IT security, physical security and records management. This will be a core course for all new managers.

To address the needs of all new employees, and not only those in the professional streams, a computer-based training module has been developed. This course will be delivered to employees at the same time that they receive their user-ID for the computer system. It takes about 20 minutes of their time and covers all of the basic information on the topics of confidentiality, privacy, network use, IT security, physical security, fire safety and building evacuation practices.

Future training plans in this area include a revision of the modules given to interviewers and courses to be developed specifically for Directors, who under the Policy on Sensitive Statistical Information have been given custodian responsibilities for statistical information holdings in their Division.

#### **4.2 Physical security**

Physical security of the Statistics Canada premises located in Ottawa and in the regions has long been the subject of considerable attention. In the mid-1980s in response to a number of false bomb-scares, the Agency imposed a higher level of physical security. Access to the premises is restricted to employees who must have an enhanced reliability security clearance. Authorized visitors to the premises must be escorted. Guards and premises detection devices monitor access twenty-four hours per day, seven days per week. After-hours access is limited to those persons authorized by their Director and such access is recorded.

The Departmental Security Office maintains the list of access privileges and ensures that the security clearances have been obtained from the RCMP. As required it undertakes preliminary investigations of any reported breach of confidentiality or misuse of government property. The

recently revised Government Security Policy also requires that threat risk assessments be conducted on a regular basis.

### 4.3 IT Security

Statistics Canada has for some time maintained a very strict code of IT security. The Agency operates two separate computer networks. Confidential statistical information is only permitted on Network A. Network A has no connection with any other network, including the Internet, thus it cannot be accessed from outside the premises. No confidential information is permitted on Network B which is connected to external services including the Internet.

Special circumstances that might require confidential information to be temporarily outside Network A are handled through explicit procedures designed for the purpose and approval by senior management. For example, field staff undertaking computer assisted interviewing do so with encrypted laptop computers and transmit the encrypted completed interviews to specific drop boxes that permit access to the Statistics Canada Network B computer system. The encrypted information is moved over into Network A only when the drop box is disconnected from Network B.

Staff are required to sign three times each year that they have read and understood the Agency's Network Use Policy. This policy governs their use of the IT system and covers issues such as acceptable use of e-mail and the Internet, and virus protection.

Rapidly evolving technology represents a significant challenge to continued IT security. The wireless world in particular confronts the Agency with some difficult decisions given the Network A/B configuration of the computer system. Under current policy the connection to Network A of wireless devices having an external connection is prohibited. A review is underway to study whether any changes are required to the Agency's Network A/B structure in the face of changing technology.

Business resumption planning is also an issue that concerns the Agency. Serious viruses can cause havoc to an information-based organization. The most up-to-date detection, isolation and irradiation methods must be in place along with employee awareness and compliance with the virus prevention protocols. Statistics Canada IT staff and employees are most vigilant in this matter as data reliability, integrity and confidentiality are at risk in the event of a serious virus attack.

The events of September 11, 2001 have also caused the Agency to reconsider its hot and cold storage plans with the view to ensuring resumption of the mission critical data programs. In this area, the agency is building on the experiences gained in dealing with earlier incidents such as the 1998 Ice Storm and the Y2K threat.

### 4.4 Disclosure avoidance practices

The *Statistics Act* requires Statistics Canada to protect the confidentiality of respondents. Considerable attention is paid by the Agency to the disclosure avoidance practices used to ensure that the identity of respondents is not inadvertently disclosed in any of its outputs.

Directors are responsible for ensuring that confidentiality is protected in their statistical products. Methodologists are

assigned to these divisions to provide assistance in ensuring that published tabulations as well as custom and cost-recovery work meet strict confidentiality requirements. Particular care is required in avoiding inferential and residual disclosures. This matter is becoming a greater issue with the growth of analytic work through the Research Data Centres (see section 5.2).

The Agency has put in place a number of checks and balances that serve to assist both Directors and the methodological staff. The Micro-data Release Committee reviews for confidentiality every submission for the production of a public use micro-data file. This approach ensures that these public use files meet rigorous standards. Recently a Disclosure Review Committee was established for the review of tabular outputs. The objective of this group is to work towards the establishment of best practices across the subject matter divisions. The advent of the Research Data Centres (Section 5.2) has also caused the Agency to review its disclosure practices. The increase in the amount and complexity of analysis being undertaken has created new challenges for methodologists and necessitated that standard approaches be adopted for disclosure review of analytic outputs.

The Disclosure Control Resource Centre along with two task forces on disclosure control (business and household areas) has increased awareness of the disclosure issue and has led to greater vigilance on the part of analysts and methodologists. The measures being taken in the Research Data Centres, along with the Disclosure Review Committee, permit the Agency to advance the application of new techniques and to gain experience in the areas of complex analyses especially in the context of longitudinal files. The existence of public use sample files and section 12 data-sharing files increases the difficulty of assessing tabular outputs for disclosure potential as these become more susceptible to both inferential and residual disclosure. This may become increasingly problematic if users attempt to use their knowledge of files to identify individuals in other outputs. While this is expressly prohibited in the contracts signed by researchers, such a risk must be considered and steps taken to verify that such disclosures do not arise.

### 4.5 Waivers and the discretionary release policy

Section 17(2) of the *Statistics Act* permits the Chief Statistician in certain instances to release information that would normally be confidential. The Chief Statistician can authorize release of individual information if the person or business gives consent. Waivers are often used in a business context to permit disclosure under this section of the Act.

The Chief Statistician can also release the names and locations of businesses for research and statistical purposes. Within Statistics Canada, the Data Access and Control Services Division receives requests for discretionary disclosures of this sort and prepares the documentation and justification. The applications for disclosures are then evaluated by the Discretionary Release Review Committee before being submitted to the Chief Statistician for consideration. Such disclosures are not permitted if the

requester has not clearly indicated that the information would be used only for research and analysis.

#### 4.6 Group privacy

Recently the Agency has considered in detail the concept of group privacy. This topic came to the forefront as the result of a complaint to the federal Privacy Commissioner by an individual who was of the view that their personal information, notably income, had been disclosed by Statistics Canada to a marketing company. The federal Privacy Commissioner reviewed the complaint and ruled that it was unfounded. Nonetheless, this caused the Agency to review both its disclosure methods and the manner in which it presented its information.

The working group that had been asked by the Chief Statistician to consider the matter, concluded that there was no accepted definition of what constitutes “group privacy”. It noted that there are two elements to the notion of “group privacy”: inferential disclosure and stereotyping of groups. The working group also noted that the mandate of the Agency is to disseminate information, and its job is to produce information about ‘groups’. To withhold information selectively about particular groups could lead to attempts to control information and could be challenged under the *Access to Information Act*. The Agency does however have a responsibility to ensure that inferential disclosure does not occur, and the Disclosure Review Committee was set up in part to address this issue.

Regarding the community harm that might occur either through stereotyping or release in a sensationalist manner of information that seems to target a group or community, the Agency adopted a number of recommendations. These included additional training for analysts so that they would become sensitive to the need not to sensationalize findings in our official releases. Marketing Division agreed to more vigilantly enforce the licencing contracts signed by secondary distributors to ensure that their descriptions of data products were not misleading. Marketing division would also review brochures and marketing campaigns to ensure that readers were fully informed that only non-confidential aggregate information was being disclosed. Research in the area of users’ understanding of confidentiality messages would also be undertaken.

### 5. Controlling use of data

As described above, the *Statistics Act* allows the collection of information for statistical purposes, while the *Privacy Act* requires that respondents be informed of uses of the data, and also recognizes statistical usage as a legitimate secondary use of data originally collected for other purposes. Despite these broadly permissive provisions, Statistics Canada still has to ensure that data in its custody are used strictly for statistical purposes and in ways that respect the intent of the *Privacy Act*. In this section, we examine some processes in place to ensure that the Agency adheres to these provisions in the use of the data that it holds.

Some analytic uses of information are becoming increasingly attractive, especially with the growing interest in understanding processes, assessing outcomes and analysing

longitudinal phenomena. Record linkage is an important analytical tool in this regard. It also raises a number of important privacy issues that must be addressed by statistical agencies. The United States General Accounting Office recently published an important study on record linkage and the privacy and confidentiality issues that it involves (US GAO, 2001). The development of Research Data Centres is another means used by the Agency to foster greater use of confidential data sets within the confines of the *Statistics Act*.

#### 5.1 Record linkage

Since the mid-1980’s, Statistics Canada has had in place a Record Linkage Policy designed to achieve the objectives of protecting the privacy of individuals while at the same time permitting record linkage under certain circumstances. Record linkage can be undertaken at Statistics Canada for research and statistical purposes only. Approval for linkage will be granted if the benefits of the proposed linkage are judged to outweigh the privacy intrusion it represents.

The linkage must lead to benefits that serve the public interest. It must be clear how the proposed methodology could lead to results that could be implemented to address important public issues. The cost and feasibility of alternatives not involving record linkage must be considered. The consent of the subjects carries a lot of weight. Lacking that consent, the likely attitude of the subjects to the linkage, or their representatives, must be considered. The impact on the Agency’s future collection programs must be taken into account. The Office of the Privacy Commissioner will be consulted whenever we consider a linkage particularly sensitive or when it involves a type of linkage not normally undertaken by the Agency.

All linkages must be approved by the Agency’s Policy Committee. Linked files must be destroyed once the approved work has been completed. Summaries of all approved linkages are shown on the Statistics Canada web site.

This approach has served the Agency well for more than 15 years. By following this strict protocol the Agency has avoided negative publicity such as that encountered by another department which was accused by the federal privacy commissioner of maintaining a “Big Brother” file on Canadians. The risk to Statistics Canada of such an accusation would be significant and serious and could put the Agency’s activities in substantial peril. Transparency, strong governing procedures and an ethical position on the undertaking of record linkage has led to the sound management of this important activity. Moreover, record linkage is a very critical research tool which if managed wisely yields considerable dividends in the areas of epidemiological studies and outcome analyses, as well as the avoidance of response burden.

#### 5.2 Research Data Centres (RDC)

Following the lead of some U.S. statistical agencies, Statistics Canada has recently established Research Data Centres (RDC) located in nine Canadian universities. These centres are enclaves of Statistics Canada within which all of



Statistics Canada's requirements of security and confidentiality apply. The establishment of RDC's is a joint project with the Social Science and Humanities Research Council (SSHRC), the Canadian Foundation for Innovation (CFI), Canadian universities and Statistics Canada. Funding was provided by SSHRC and CFI for the initial start-up costs with the universities bearing the continuing costs. Statistics Canada supplies the confidential data files, ensures that disclosure avoidance is undertaken for all outputs from the RDC, and participates in the peer and institutional approval processes for the projects and the final products.

Access to the confidential data files is permitted under section 5 of the *Statistics Act* which permits 'deemed employees' to perform a service for the Agency. All proposed research projects are reviewed by a committee comprised of representatives from SSHRC and Statistics Canada who determine the scholarly merit of the work and verify that the work can only be undertaken with access to the confidential data files. In cases where the public use files would be appropriate or the work lacks rigour or focus, the application will be denied. Approved researchers have access to the required files within a RDC, but only results that have been screened for disclosure protection can be taken outside the RDC. Researchers are required to produce a report for Statistics Canada that it may publish. Once that obligation is fulfilled, researchers are free to publish other articles that may be based on the research project.

These centres have significantly increased the access by researchers to the complex longitudinal survey files. These are files from which it has become impossible, due to risk of disclosure, to produce useful longitudinal public use files. In the first year of operation more than 125 projects were initiated in these centres. The research papers that will be produced and published by Statistics Canada will significantly contribute to our understanding of a number of important social policy issues in the areas of health, transitions from school to work, child development, labour and employment dynamics and the workplace.

### 5.3 Archiving

The final step of data stewardship is that of archiving the final master file. The *National Archives Act* along with the federal government policy on the Managing of Government Information Holdings (MGIH) requires that government institutions fully protect and safeguard for future generations those government records deemed by the National Archivist as warranting preservation. In the statistical program, a Record Disposal Authority is required before a survey master file can be destroyed.

Statistics Canada initiated in the mid-1980's an archival system for final master files which has subsequently been incorporated into the Agency's Integrated Meta Data Base (IMDB). Among other data elements, this data warehouse contains descriptive information concerning the location, format and content of the confidential master files for all of Statistics Canada's surveys. The confidential master files and all electronic meta data elements are preserved and routinely exercised to ensure their continuing accessibility.

## 6. Privacy Impact Assessments

Beginning in May 2002, Statistics Canada has been required to implement a new Government policy (Treasury Board of Canada, 2002) requiring that all federal institutions develop and maintain Privacy Impact Assessments (PIAs) to evaluate whether program and service delivery initiatives involving the collection, use or disclosure of personal information comply with privacy requirements and resolve privacy issues. The Agency response to this policy has been to develop and maintain generic privacy impact assessments. These will address the majority of the Agency's survey and administrative data collections, uses and disclosures. Where the Chief Statistician decides that a particular survey or administrative data collection, use or disclosure would require a unique PIA, then the generic materials would be used as the basis for the identification of the specific privacy concerns.

The Agency has developed a generic response to the 10 privacy principles identified in the PIA policy.<sup>2</sup> To each of these principles, the Agency has identified the policies, procedures, guidelines and legal authorities in place to address them.

Regarding the PIA policy requirement to undertake a data risk assessment and mitigation analysis, again a generic approach has been used. The risk assessment and mitigation measures in place for the commonly used data collection methodologies have been prepared.

Thus, specific and unique privacy issues need to be identified only for those surveys that the Chief Statistician deems as requiring a separate PIA. This approach will reduce the burden of implementing the PIA policy, yet at the same time serve to demonstrate to Canadians that all surveys and administrative data uses conducted by Statistics Canada are undertaken in a manner that is mindful of privacy implications. Moreover this is an approach that has been adopted by other organizations, for example the Canadian Institute for Health Information has recently published their response to the ten privacy principles (see [www.cihi.ca](http://www.cihi.ca)).

## 7. Conclusion

Respect for the privacy of respondents and the maintenance of the confidentiality of the individual responses are survival issues for Statistics Canada. This paper has attempted to describe the approaches used in Statistics Canada for managing privacy and confidentiality issues.

A key element in managing privacy is maintaining a productive and constructive relationship with the Office of the Privacy Commissioner. This has to date been achieved by ensuring that an open dialogue is maintained with the Privacy Commissioner and his staff regarding any activities that have a privacy impact. For example, surveys likely to have privacy implications are discussed with the staff of the Office of the Privacy Commissioner. Regular briefings of

---

<sup>2</sup> These ten principles are: accountability, collection, consent, use, disclosure, accuracy, safeguards, openness and transparency, individual access, and challenging compliance.

the provincial counterparts are undertaken by the Regional Directors of Statistics Canada. Program area survey managers may also be involved in these briefing sessions, both at the federal and provincial/territorial levels.

With rapid advances in electronic communication technology, and growing awareness of the privacy concerns that this technology brings, we can expect that statistical agencies will find themselves under increasing pressure to justify their activities in the context of privacy. An open and coherent approach to addressing privacy issues on the part of statistical agencies will be a necessary element of their successful management in the future.

#### References:

- Canada, *Privacy Act* (1985), R.S. 1985, c. P-21.
- Canada, *Access to Information Act* (1985), R.S. 1985, c. A-1.
- Canada, *Statistics Act* (1985), R.S. 1985, c. S-19.
- Rancourt, E. and Hidirolou, M. (1998). *Use of Administrative Records in the Canadian Survey of Employment, Payrolls and Hours*, "Proceedings of The Survey Section of the Statistical Society of Canada, 39-49.
- Särndal, C.-E., Swensson, B., and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New-York: Springer-Verlag.
- Statistics Canada (1988). Policy on the Security of Sensitive Statistical Information, February 1988. Policy Manual 4.7.
- Statistics Canada (1996). Policy on Record Linkage, July 1996. Policy Manual 4.1
- Statistics Canada (1997). The Determination of Mandatory and Voluntary Surveys: Guidelines. Internal Document.
- Statistics Canada (1998). Policy on Informing Survey Respondents, September 1998. Policy Manual 1.1
- Statistics Canada (2002). Policy on the Review and Testing of Questionnaires, October 2002. Policy Manual 2.8
- Treasury Board of Canada (2002). Privacy Impact Assessment Policy, May 2002.
- United States General Accounting Office (2001). "Record Linkage and Privacy. Issues in Creating New Federal Research and Statistical Information", April 2001.