Culture, Tourism and the Centre for Education Statistics

Postsecondary Student Information System (PSIS): An Overview

August 2002

Statistics Canada Statistique Canada

Canada

Table of Contents

1 INTRODUCTION

Statistics Canada is developing the Postsecondary Student Information System under the aegis of the Canadian Education Statistics Council. The Council is a joint body of Statistics Canada and provincial and territorial deputy ministers responsible for education. It is chaired jointly by the Chief Statistician of Canada and a Deputy Minister. The individual departments act collectively through the Council of Ministers of Education, Canada.

The primary objective of PSIS is to meet policy and planning needs in the field of postsecondary education and the transition to the labour market.

Canada's place in the global economy depends to a high degree upon the knowledge and skills of its citizens. The postsecondary education system is an important vehicle for delivering the required knowledge and skills. For Canada to compete in the new economy informed program choices by students and effective marshalling of education resources are essential.

This report provides an overview of PSIS and the nature of the information it will provide. It addresses questions concerning privacy, confidentiality and the security of data. These questions are also examined in more detail in a companion document, *Postsecondary Student Information System: Addressing Privacy Concerns*.

PSIS is designed to hold a complete inventory of all Canadian postsecondary institutions and the programs and courses they offer, as well as demographic, program and course information for each student registered at these institutions.

PSIS replaces the current administrative surveys of postsecondary enrolment and graduation. The University Student Information System (USIS), the Community College Student Information System (CCSIS) and the Trade and Vocational Survey (TVOC) have been in place for many years and were in need of updating. PSIS addresses shortcomings of these surveys and provides additional information to address the policy and planning needs of the Canadian Education Statistics Council and other education stakeholders.

2 BACKGROUND: HOW PSIS CAME ABOUT

Statistics Canada began compiling and publishing postsecondary enrolment and graduation data in the 1920s. Since the early 1970s, under the authority of the *Statistics Act*, the data have been drawn from student records submitted by universities and colleges in Canada. The data come from the administrative systems of postsecondary institutions. In some cases, institutions submit their

data directly to Statistics Canada. In other cases, data are submitted to provincial ministries or agencies, which in turn report to Statistics Canada.

Over the years, these data have shaped our understanding of the postsecondary system and are important nationally and internationally. Nationally, the data feed annual provincial comparisons and trends. Combined with information on postsecondary finance and human resources, these data provide important information about the efficiency and effectiveness of the postsecondary system. Internationally, the data position Canada in comparisons undertaken by the Organisation for Economic Co-operation and Development (OECD) and the United Nations Education, Scientific and Cultural Organisation (UNESCO). Increasingly, the data are important to Canada's position in free trade discussions.

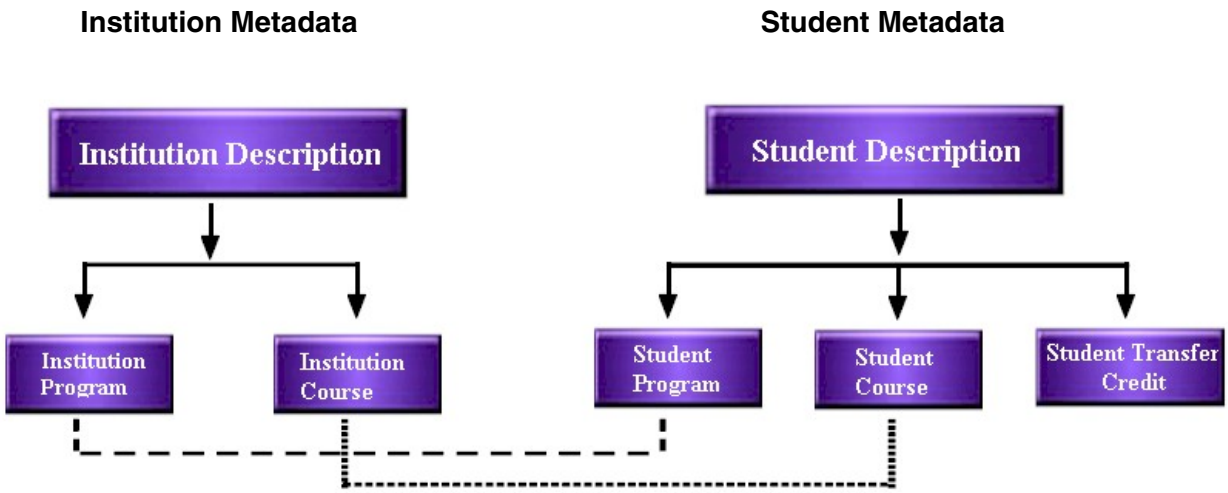
In recent years, however, the public, researchers, provincial ministries and postsecondary institutions have voiced demands for more and better data to address increasingly complex questions. In 1997, the Council of Ministers of Education (CMEC) and the Centre for Education Statistics at Statistics Canada undertook a joint strategic planning exercise. The purpose was to identify education policy and management issues and the information needed to address them. All provincial and territorial ministries responsible for education and over 100 national and provincial associations were consulted. PSIS emerged as a lynchpin in the resulting Strategic Plan for the Centre for Education Statistics. The Canadian Education Statistics Council (CESC) approved the Strategic Plan in 1997.

Statistics Canada provided initial funding for the PSIS project. Two pilot projects, one involving Maritimes universities and the other involving Ontario colleges, served to define data requirements and to assess feasibility. Consultations with governments and institutional data providers further refined the data requirements. Based on the positive results of the pilot and the consultations, the federal Policy Research Initiative (PRI) provided funding for implementation. The objective of the PRI is to support research that will contribute to policy development and initiatives to improve Canada's competitive position in the global economy. The PRI recognised the role PSIS could play in increasing the effectiveness of Canada's postsecondary education system. More recently, other federal government departments have shown interest in the work. Human Resources Development Canada (HRDC) has been willing to contribute financially because PSIS will be able to deliver better planning data and because it will be able to provide a comprehensive, accurate, consolidated list of postsecondary institutions and the programs they offer. This comprehensive view and accurate registry of institutions and programs is essential to the work of the Canada Student Loans Program managed by HRDC.

3 CONTENT OF THE PSIS DATABASE

The PSIS database holds data at two levels: (i) institution and program data, and (ii) student data. The institutional and program data includes a complete list of all postsecondary institutions in Canada, and a complete inventory of all programs and courses offered through these institutions. The student data contains demographic, program, and course information for students registered at those institutions. As shown in the accompanying figure and as described in sections 4 and 5, the PSIS database is fed by seven input files that describe either the student or the institution and the programs it offers.

PSIS Input Files



PSIS data is reported once a year, either directly by institutions or indirectly by co-ordinating organizations reporting on their behalf. The PSIS final report covering the full academic year occurs in July. The final data will give a complete picture of academic activity over a full 12 months.

4 INSTITUTION AND PROGRAM DATA

Three files describe the institution's programs and courses:

Institution Description File: Lists and describes the different periods of academic activity by which programs and courses are organised.

Institution Program File: Program name, duration, credential, prerequisites and other characteristics of each program offered by the institution.

Institution Course File: Course name, duration, credit value and other characteristics of each course offered by the institution.

These files consolidate information that is already on the public record, in either electronic or paper form. The names of institutions and the programs they offer can be found in individual institutions' publications, in government publications produced for student counselling purposes, in private sector publications, or in the publications of provincial or national organizations such as the Association of Universities and Colleges of Canada (AUCC) and the Association of Canadian Community Colleges (ACCC). However, the existing databases all have limitations in terms of completeness, coverage, and lack of standard coding of the program information. PSIS will overcome these limitations by consolidating the information into a single harmonized source for the entire postsecondary system.

This comprehensive and accurate registry of institutions and programs is essential to the work of Statistics Canada in planning and carrying out the postsecondary education statistical program.

5 HOW PSIS WILL IMPROVE INFORMATION ON INSTITUTIONS AND PROGRAMS

a) PSIS will provide complete and accurate information on programs offered by postsecondary institutions, coded to the Classification of Instructional Programs (CIP). No such comprehensive database of postsecondary institutions and coded inventory of programs currently exist. The institution lists currently maintained by the Centre for Education Statistics are incomplete, and systems used for classification of education programs are not comparable across different levels of postsecondary education – university versus college and trade vocational. PSIS will permit dissemination of comparable program information by level of education.

b) Because CIP is being used, PSIS will make international comparisons possible. CIP has been accepted as the statistical standard on program classification by the U.S. Department of Education's National Center for Education Statistics and Statistics Canada.

c) The PSIS institution and program database will provide frame information for the National Graduates Survey (NGS). The design of the NGS sample requires a list of institutions and the number of graduates by program. In past NGS surveys the lack of comparability of the program information maintained by Statistics Canada across different levels of post education has been problematic.

d) Subject to appropriate data-sharing arrangements[1] under the *Statistics Act*, Statistics Canada would be able to share institution and program information with other organizations having a requirement for such information. This will reduce the respondent burden on institutions, who currently have to provide such information to multiple organizations; it will eliminate the inefficiency of maintaining separate lists; and it will overcome the data quality problems associated with existing lists.

6 STUDENT DATA

Four files provide descriptive information about students and each program and course students were enrolled in during the reporting cycle. Unlike the institution and program files, these files contain data on individual students.

Student Description File: Student name, birth date, gender, social insurance number (SIN), personal contact information, previous education, and characteristics such as whether or not the student has self-identified as a visible minority, aboriginal person or person with a disability. There is one record per student per institution. This file also contains the PSIS National Student Number (PSIS-NSN). The PSIS-NSN is a unique student identifier designed to help link incoming student records with records already on the national database. The PSIS-NSN is described later in this paper.

Student Program File: Contains one record for each program in which the student was enrolled during the reporting cycle. The student program record includes when the student started/ended a program, student status (still enrolled, completed/graduated, withdrew, did not successfully complete, other), majors, transfer credits, fees billed, credits earned to date and other characteristics of the student's program as recorded by the institution.

Student Course File: Contains one record for each course in which the student was enrolled during the reporting cycle. The student course record includes when the student started/ended the course, student status (still enrolled,

[1] Section 12 of the *Statistics Act* makes these specific provisions for Data Sharing agreements:

> (Statistics Canada) may enter into an agreement with any department or municipal or other corporation for the sharing of information collected from a respondent by either Statistics Canada by the department or corporation on behalf of both of them and for the subsequent tabulation or publication based on that information. An agreement shall provide that
> (a) the respondent be informed by notice that the information is being collected on behalf of Statistics Canada and the department or corporation, as the case may be; and
> (b) where the respondent gives notice in writing to the Chief Statistician that the respondent objects to the sharing of the information by Statistics Canada, the information not be shared with the department or corporation…

completed/graduated, withdrew, did not successfully complete, audited, other), their credits earned, fees billed and other characteristics of the student's course as recorded by the institution.

Student Transfer Credit File: Contains one record for each course credit the student was granted for courses taken and/or experiences outside the current institution. To the extent that this information is stored in the administrative system of the institution, the transfer course record contains the number of transfer credits, the source of the credit and the date it was granted.

7 HOW PSIS WILL IMPROVE INFORMATION ON STUDENTS

PSIS student data will replace the current administrative surveys of postsecondary enrolment and graduation: the University Student Information System (USIS), the Community College Student Information System (CCSIS) and the Trade and Vocational Survey (TVOC). Some examples of the ways in which PSIS will improve information for policy needs are described below:

a) The existing college and university surveys are "snapshot" surveys. Enrolment counts are taken at one point in time during the year. As a result, current surveys miss a significant amount of enrolment information. PSIS will capture all academic activity over a twelve-month period, and thus provide a comprehensive understanding.

b) Existing surveys capture only major field of study. In contrast, PSIS will reveal how programs and student course mix are evolving in response to labour market and other demands.

c) PSIS will capture information on student enrolments at a course level, including information about the duration and credit value of courses. This will yield a precise, standard measure of student workload to replace the existing imprecise measure known as 'full time equivalent' or 'fte' This is a variable that requires accurate measurement. It has implications for the length and costs of studies, for persistence and for the capacity of the postsecondary system.

d) Existing surveys do not provide information about foreign students pursuing their studies through Canadian institutions, either at a campus in Canada, at a campus overseas, or at their own home or workplace through distance education. PSIS will provide data on all such students. In the global environment, governments are increasingly interested in supporting and evaluating the impact of educational experience undertaken outside Canada.

e) PSIS contains new data elements such as program prerequisites, co-op education, scholarships and bursaries, courses delivered under contract, program capacity, and mode of course delivery. Further, PSIS is designed to

accommodate future changes to the structure of postsecondary education and the emergence of new information needs.

f) Life-long learning - PSIS will provide information about participation in postsecondary education later in life. In combination with sample surveys and other databases, the impact of continued learning on career and income (productivity) can be measured

g) Effectiveness of the education system - Governments are concerned about the outputs of the education system relative to the resources spent. Measures such as number of completions, time required to complete, persistence and retention will provide measures of the relationship between educational inputs and outcomes achieved. PSIS will improve the ability to make these measures standardized and comparable.

h) Transferability of credits - Increasingly students are moving between fields of study, institutions and levels (i.e., between universities and colleges), to construct their desired package of knowledge, skills and credentials. The extent to which transfer credits are granted by institutions has a strong impact on the efficient use of educational dollars. PSIS will provide comparative measures related to the granting of transfer credits.

i) Co-operative program delivery - Many institutions, often colleges and universities, are offering a range of programs variously known as articulated, collaborative and brokered. Articulated programs may for example begin with college level nursing and continue with university level nursing science with full recognition given to the college level program. Collaborative programs may combine college level courses with university level courses leading to an applied degree. In brokered programs institutions arrange to have their students take some of their courses at a second or third institution. PSIS will provide a measure of the increase in such program delivery activity.

j) Partnering with industry - Many institutions today are partnering with the private sector to deliver such education services as tailored degrees and employee training. PSIS will provide a measure of the extent and nature of these activities.

k) Student mobility - Provincial governments are interested in the mobility of students. Some provinces experience significant out-migration while the reverse is true for others. PSIS will allow for the study of student flows.

l) Student pathways - PSIS will trace pathways as students move between fields of study, between institutions, between university and college and between full and part-time status.

m) Progress of the secondary population - Provinces are also interested in the future progress of their secondary populations. PSIS will provide feedback to provinces about the percentage of students who go on to postsecondary education anywhere in Canada, at what level and with what result.

8 POTENTIAL FOR ENRICHING THE INFORMATION THROUGH LINKS TO OTHER SOURCES OF INFORMATION

By linking the information obtained from institutions to other sources of statistical information at the micro level, it would be possible to address important policy and planning needs. In the future, respondents to voluntary sample surveys, such as the Youth in Transition Survey, will be asked for their consent to combine the data they provide in their survey responses with the data contained in PSIS.

A Note on Record Linkage

Statistics Canada collects information from a variety of sources. The Centre for Education Statistics itself has data from sample surveys such as the Youth in Transition Survey and the National Graduate Survey, as well as data from administrative surveys such as PSIS.

Under very strictly controlled conditions, and only for research or statistical purposes, data from two or more surveys can be linked within Statistics Canada. When two or more sources are combined in this way, the 'link' is limited to a predetermined period of time and for very specific analytical purposes.

The strict controls on record linkage are found in Statistics Canada's policy on record linkage. The Chief Statistician must approve all proposals for record linkage, and a description of all linked files will be available on Statistics Canada's web site.

Similarly, the potential exists to add information on student loans or income after graduation through linkages to administrative files. Any such linkages would be subject to prior approval under the Statistics Canada Policy on Record Linkage. The linked files would be on a sample basis only, without personal identifiers, used only for approved research and statistical uses, and retained only for the duration of these studies.

The sample linkage strategy described above would be greatly facilitated by the use of SIN as a common and unique identifier between PSIS and other files being linked. The linkage strategy could involve selection of a sample of records from

PSIS to be representative of different programs and levels of study within provinces or territories. The linkage would occur only among those records on the other files with SIN matching those of the sample of PSIS records. Without a unique identifier, it would not be possible to employ such sampling strategies involving samples of both administrative data sets.

These are examples of the benefits that might be gained by linking PSIS to other data sources:

a) Youth in Transition Survey (YITS):
Combining data from PSIS and YITS would offer insights into postsecondary access, retention, persistence, time to completion and outcomes in light of youth characteristics, behaviour, attitudes and aspirations, family background and experiences during secondary and postsecondary schooling.

b) National Graduate Survey (NGS):
Combining PSIS and NGS would provide much more detailed information about programs and courses than is currently available in NGS, to better understand knowledge and skills acquired in relation to labour market outcomes. It would shed light on the relationship between mobility and pathways, and labour market outcomes – without adding response burden or data collection costs.

c) Harmonised Student Loans Program (HSLP):
Linking HSLP data with PSIS and NGS data would help to understand the impact of the loans program on student persistence and education and labour market outcomes. It should be noted that administration of the HSLP program is the responsibility of Human Resources Development Canada, and the HLSP database resides at HRDC. In addition to satisfying the requirements of Statistics Canada's record linkage policy, such linkages for statistical and research purposes would only occur if HRDC agreed to the linkage and provided Statistics Canada extracts of the HLSP database that could be used for this purpose.

e) Income Tax file:
At the present time, knowledge about the long-term earnings of graduates from different disciplines is incomplete. Surveys such as the Labour Force Survey and Census provide approximate information. In order to gain more in-depth understanding of this issue, a linkage of USIS with the tax information provided to Statistics Canada by Canadian Customs and Revenue Agency (CCRA) under the provisions of the *Statistics Act*, has been approved under Statistics Canada's record linkage policy. The existing study, while yielding very useful results, nevertheless has limitations. SIN is only reported at the university level in current data systems and is missing for a number of institutions, including all institutions in Quebec. With more complete reporting of SIN under PSIS, it would be possible to provide a complete and accurate view of the

long-term earnings across all levels of postsecondary education of graduates from different disciplines.

9 CONFIDENTIALITY, PRIVACY AND THE SECURITY OF INFORMATION

All surveys of postsecondary enrolment and graduation undertaken by the Centre for Education Statistics collect individual student record data. However, unlike its predecessors, PSIS asks for student identification data such as name and address. The inclusion of student identification data raises privacy issues.

Student identification data are included for three reasons. First, they provide a means to contact respondents in sample surveys. Second, they provide the means to link data over time. Third, they will ensure accuracy.

Statistical vs Administrative

Statistics Canada's purpose is statistical rather than administrative. Unlike databases having to do with such administrative functions as motor vehicle licensing which directly affect the individuals involved, the student data held in PSIS, as well as any statistical files created through linkage, have no administrative purpose.

No decision will be made about any individual based on the information in Statistics Canada's data files. The files are used to produce analyses of trends and patterns. Groupings rather than individuals are the focus of analysis and Statistics Canada ensures that individuals cannot be identified in the resulting tables and graphs.

Contact: Statistics Canada undertakes follow-up surveys of students and graduates. The NGS is an example that is well known in postsecondary education. To do such surveys, in the past Statistics Canada has selected a sample, using the unique identifiers assigned by colleges and universities and included on the individual student records. Statistics Canada has then asked institutions to provide the corresponding name and contact information.

Both Statistics Canada and institutions have found this process expensive and time-consuming. Under PSIS the name and contact information needed to carry out these surveys will be submitted as part of each student record.

Linkage: A fundamental objective of PSIS to provide information on student mobility across institutions or provinces over an academic career that may cover

many years. Unlike the existing surveys in which individual records submitted from year to year are not linked, each student record in PSIS will be longitudinally linked to represent a cumulative history of the student's progress through the postsecondary system. This measure provides for the study of mobility, pathways and outcomes as described earlier.

Accuracy : PSIS asks for a number of student identification elements to ensure a high degree of accuracy. The need for accuracy becomes especially acute because of the fundamental need for PSIS to provide information on student mobility over more than two or three consecutive years.

The inclusion of personal identifiers in PSIS has triggered questions about Statistics Canada's authority to obtain the information, and about confidentiality, privacy and the security of information. This part of the report addresses these concerns and related matters, such as record linkage, informing students about uses of the information, the purposes of the PSIS National Student Number and the Social Insurance Number, and retention of records.

Statutory Authority for PSIS

The *Statistics Act* mandates Statistics Canada to carry out a program of education statistics:

- Section 3 mandates Statistics Canada "to collect, compile, analyse, abstract and publish statistical information relating to the commercial, industrial, financial, social, economic and general activities and condition of the people."
- Section 22 directs Statistics Canada to "collect, compile, analyse, abstract and publish" statistics on education.
- Section 13 sets out the legal authority for Statistics Canada to obtain access to "any documents and records that are maintained in any department or in any municipal office, corporation, business or organization", where Statistics Canada requires those records for the purposes of the *Act*.

Confidentiality of information

To balance Statistics Canada's powers to collect information, the Agency is under rigorous legal obligations to protect the confidentiality of information about individuals:

- Paragraph 17(1)(b) of the *Statistics Act* prohibits the disclosure of any information obtained under the *Act* "in such a manner that it is possible from the disclosure to relate the particulars obtained from any individual return to any identifiable person, business or organization".
- The *Statistics Act* also states that all Statistics Canada employees must take an oath not to disclose any confidential information related to any person, business or organisation. Contravention of the oath resulting in a violation of confidentiality is a **criminal offence** punishable by fine and/or a term of imprisonment.

- Section 24 of the Federal *Access to Information Act* prohibits the head of a government institution from disclosing any record that is requested under the *Act* if it contains information whose disclosure is restricted by certain provisions in other statutes. Section 17 is one such provision; information that cannot be disclosed under section 17 of the *Statistics Act* also cannot be disclosed under the *Access to Information Act*.

All Statistics Canada employees having access to confidential statistical information must undergo a security check. Only Statistics Canada employees who need to view confidential files as part of their duties are authorised to access them. In reality, access to confidential is limited to the very small number of staff who either maintain databases or who undertake analysis and research. In the case of linked files, the number is restricted to those working on the research project that requires the linked file actually have access.

Maintaining the security of PSIS data transmissions

PSIS data will be transmitted to Statistics Canada over the Internet, using secure, government-approved encryption procedures. Currently, respondents use software developed by Statistics Canada known as the *Data Return Facility.* The software encrypts the files and sends them to a specified Internet address (so there can be no error in destination).

Statistics Canada is also piloting the use of PKI (Public Key Infrastructure) as a means of transmitting individual record data back to respondents for data correction and verification. PKI is a secure transmission protocol widely used in banking and other industries. PKI uses the very powerful Entrust encryption software. PKI would be used only to return confidential data to the data provider for initial verification and later to provide a copy of the provider's final data once processing and verification is complete. When data are transmitted using PKI, it would only be sent to a specific computer and be accessible only by the authorised person who holds the "key". The use of PKI would contribute to lowering costs while improving both data quality and timeliness.

Security of information

All Statistics Canada employees are responsible to ensure the security of information. Statistics Canada has in place a network of physical security systems and procedures to protect confidential information against unauthorised access. Some important features are as follows:

- controlled entry systems are in place to provide against unauthorised access in all Statistics Canada buildings;
- policies prohibit staff from removing confidential statistical information – either hard copy or electronic - from the secure environment of Statistics Canada’s workplace;
- confidential statistical information is stored in a secure fashion; and,

- disposal of confidential statistical information is carried out under secure conditions.

To ensure the appropriate protection of its information and data holdings, Statistics Canada maintains separate internal and external network environments, Network A and Network B. While both networks are protected by a number of hardware and software features, as an additional measure of protection, data which are collected under the *Statistics Act* are only maintained in the Network A environment. This network has no direct connection to the Internet, ensuring there can be no outside access of an unwanted or unintended nature. Once the encrypted PSIS information has been transmitted from the respondent to Statistics Canada's Network B environment, it will be physically transferred to the Network A environment before it is unencrypted, with none of the transferred information remaining in the Network B environment.

Maintaining confidentiality in statistical outputs

The main PSIS database is housed on its own secure server in a locked room and only a limited number of Statistics Canada employees have access to it. Statistics Canada analysts have access only to a separate research database that is stripped of personal identifiers. The software used for tabulation employs automated random rounding and cell suppression procedures.[2] Even so, there are additional procedures in place:

- All statistical outputs are visually screened to ensure that sensitive information about individuals cannot be identified. This means that an analyst at Statistics Canada manually reviews all tables before they are released.
- There are procedures to prevent residual disclosure. Residual disclosure refers to a situation in which two separate data files, if they were brought together, could result in the identification of an individual.

These procedures have been rigorously employed for years with the existing individual level university and college databases.

Record linkage

Record linkage is recognised as a valid statistical tool and as an alternative to asking respondents to provide data many times. Nonetheless, because the practice may raise privacy concerns, since 1986 Statistics Canada has in place a Policy on Record Linkage. This policy seeks to ensure that a balance is maintained between the public good to be served by record linkage and the potential privacy issues. Statistics Canada policy and procedures ensure that records will be linked only if **all** the following conditions are satisfied:

- the purpose of the record linkage activity is consistent with the mandate of Statistics Canada as described in the *Statistics Act*; and

[2] Random rounding means that all frequencies between one and 9 are rounded to 0 or 5, and this is done in a random fashion. Cell suppression means that all cells with a frequency under five (after rounding) are not visible to the user.

- the products of the record linkage activity will be released only in accordance with the confidentiality provisions of the *Statistics Act* and with any applicable requirements of the *Privacy Act*; and
- the record linkage activity has demonstrable cost or respondent burden savings over other alternatives, or is the only feasible option; and
- the benefits to be derived from such a linkage are clearly in the public interest; and
- the record linkage is judged not to jeopardise the future conduct of Statistics Canada's programs; and
- the linkage satisfies a prescribed review and approval process set out and managed by Statistics Canada.

Further, where any record linkage activity is proposed that involves linking existing records to information Statistics Canada would collect directly from respondents, they must be notified of the proposed linkage activity at the time of collection. An example would be a linkage between PSIS and NGS. Respondents would be asked as part of NGS if they consent to linking their response with their student record in PSIS. They would be told the purpose of the linkage and the value of the resulting information.

Informing students of the uses of their information

In keeping with the privacy principle of informed consent, Statistics Canada will be working with universities and colleges to develop means for institutions to inform students about the way in which the information will be used and their rights with respect to that use. The material will also make it clear that their information is provided to Statistics Canada under the authority of the *Statistics Act*, for statistical and research purposes. Institutions sometimes do this by including a statement to that effect in their calendars. Appendix A provides the standard wording that Statistics Canada is suggesting to institutions.

Although the surveys that PSIS is replacing also gathered individual student records with unique identifiers, the addition of name and contact information combined with increasing public concern generally about privacy and confidentiality, has prompted Statistics Canada to ask institutions to provide more information to students, as described above. In particular, institutions are being encouraged to include reference to the PSIS web-site at Statistics Canada where students can learn more about PSIS and the uses of PSIS data, and make contact with Statistics Canada should they have further questions.

From a legal perspective, when information is obtained from administrative records under the authority of the *Statistics Act*, there is no right of refusal provided to individuals to whom the records pertain. In practice, when asked to do so, Statistics Canada will delete individuals' identifying information (name, address, and SIN) from the PSIS database. We need to take steps to ensure that students are notified that this is an option, and we are seeking institutions' co-

operation in making this known. A brief note on institutions' registration forms and in their calendars is one clear way in which this can be done; wording is suggested in the Appendix.

The PSIS National Student Number (PSIS-NSN)

The PSIS-NSN is assigned to allow *probablistic* identification sufficient to make research possible, rather than the *certain* identification of individuals that is needed for a number to be used for administrative purposes. The PSIS-NSN will help make it possible to link the information institutions submit each year with the student information already held on the national database.

The capacity of PSIS to provide for the analysis of mobility, pathways and their relationship to outcomes is entirely dependent upon maintaining the uniqueness of student records - only one record for each student. Linking student records within the national database is the sole purpose of the NSN. It has no administrative purpose. Unlike a health number that must be presented to obtain health services, students will not use the number to obtain education services. In fact, students do not have to know or remember the national number. The number is assigned by the institution according to a formula provided by Statistics Canada, and carried on the institution's student record system. When the institution reports to Statistics Canada the number is part of the PSIS record. When a student moves to another postsecondary institution the number is carried on the student transcript. The receiving institution captures the number from the transcript onto its student record system. When the second institution sends its report to Statistics Canada, the NSN will be used to make the link with the existing record for that student. Statistics Canada will have the only complete NSN file and the NSN will not appear in any information output.

The Social Insurance Number (SIN)

Just like the enrolment surveys that it is replacing, PSIS includes the Social Insurance Number (SIN)[3]. One of the key features of PSIS will be its longitudinal nature. A longitudinal base of student data is needed in order to understand student mobility and transitions: in and out of postsecondary education, between institutions and provinces, and across different sectors of postsecondary education. The SIN will be used in conjunction with other variables to link student records over time. That is, while institutions and co-ordinating agencies will be reporting annual "snapshot" data on students, these student records will be linked over time in the PSIS database, through use of SIN and other variables.

In order to study student flows in and out of postsecondary education, and between institutions and provinces, records have to be linked with very low error rates.

[3] Approximately 70% of the university records outside Quebec contain the SIN.

Name and demographic variables (such as date of birth, and marital status) would not provide enough accuracy. Experience at Statistics Canada has shown that those variables allow matches up to 80% in two consecutive years. However, over time this rate would become too low for the data to be useful – especially over the decades that are implicit in the concept of 'lifelong learning'. Thus, the need for SIN relates to the accuracy principle; that is, personal information needs to be as accurate, complete, and up-to-date as is necessary for the purposes for which it is to be used.

The Social Insurance Number would also make it possible to accurately link a sample of students from the PSIS database with a sample from either the student loans database or a sample from the tax files. Such linkages would make it possible, for example, to examine the distribution of incomes by field of study 10-20 years after graduation. As described earlier, SIN is needed to accurately link these files, and the presence of SIN as a common identifier on both files is a key to the planned strategy of linking these files on a sample rather than a 100% basis.

Retention of Records

Statistics Canada regularly reviews the need to retain records, and records are regularly destroyed. In the case of PSIS, this review will need to consider the need for PSIS data to deal with the increasing policy interest in the notion of life-long learning. The policy questions presume some ability to track the education of individuals throughout their lives, even after long breaks in their educational career.

Access by individuals to their own records

Under the federal *Privacy Act*, individuals can request access to their own, individual information held on federal information banks, including those held by Statistics Canada. Existing student databases do not contain names, so that such records can be accessed only if the individual knows the national student number(s) assigned by the institution(s) he or she attended. PSIS will contain the individual's name and can be accessed directly using the name.

Access to PSIS data for analysis

Access at Statistics Canada: The main PSIS database holds name, contact information and unique identifiers. Access to this database is limited to Statistics Canada employees who have undergone an enhanced reliability check and have received specific authorization to work directly with the data. This small number of employees are responsible for maintaining the data and for ensuring that information taken from the database will not allow individuals to be identified. As mentioned earlier, security precautions include keeping the database on a secure server in a locked room.

Statistics Canada researchers do not have direct access to the main PSIS database. Their research is undertaken on research databases that do not hold personal identification information.

To ensure that no individual can be identified in statistical outputs, the software used to query the base automatically invokes random rounding and cell suppression. In addition, outputs are checked visually by analysts with subject matter expertise. In practice, this means that every table is examined so that expert human judgement ensures that individuals can not be identified.

Access by data providers: Data providers (institutions and in some cases governments or agencies) will have their own student data with additional derived variables returned to them. This is done so institutions have a copy of the complete set of data Statistics Canada holds on their institution. Using this data and software STC will provide to each institution, the institution will be able to undertake the same analysis on their own data that STC is able to undertake on national data. Data providers will be able to use their own data and the software to develop queries for comparison of themselves with non-confidential results for other data providers. Such queries can be e-mailed to Statistics Canada and run against the national research database. Outputs will be visually checked to ensure confidentiality and returned to the requestor.

Access by others: Governments, organizations, independent researchers and members of the general public that do not provide data but who need access to the data will be provided with a mock database[4] and software with which to develop their queries. Similarly, their queries can be e-mailed to Statistics Canada where they will be run against the research database and vetted for confidentiality before being returned to the requestor.

Consideration will also be given to development of a public use microdata file. 'Public use microdata files' are files of individual records that have been stripped of identifiers and reviewed by an internal Statistics Canada technical committee to ensure that the records do not contain any variables in a format that would allow the re-identification of any respondent.

A companion report, *Postsecondary Student Information System: Addressing Privacy Concerns* details Statistics Canada's commitment in the light of existing national privacy standards.

[4] A "mock" database contains "pretend" data that looks similar to the real data. The "mock" and "real" databases have the same structure so that a client is able to generate database queries that can be run against the "real" research database by Statistics Canada.

10 PRODUCTS AND SERVICES

Statistics Canada will publish PSIS data in the form of standard statistical tables and research papers. Standard tables will be published in Education in Canada, and will be available throughout the year from the Centre for Education Statistics. Research papers are published in Education Quarterly Review. PSIS data will also be used to produce a number of the indicators now published in Education Indicators in Canada as well as the annual OECD publication Education at a Glance. Statistics Canada publications are available through subscription and are also made available to the public at no charge through the federal government depository library program.

Work is also well advanced to improve the accessibility of data to institutions and other respondents (such as Ministries of Education) that supply data. These groups will receive their own data plus derived variables along with a query tool developed by Statistics Canada.

Using the query tool and their own data, they will develop queries. The queries will be e-mailed to Statistics Canada, run against the PSIS research base, and returned. The research base will be stripped of personal identifiers and programmed to invoke random rounding and cell suppression to protect confidentiality. Outputs will also be visually checked to ensure confidentiality.

Users of PSIS who are not data providers will be given a mock database and query tool. With these, they will also be able to develop their own queries to be run against the research base at Statistics Canada.

Individual academic researchers and various organizations with a need for data to develop policy – organizations such as the AUCC and the ACCC – may also contact Statistics Canada to obtain tailored data outputs and analysis. These requests are dealt with individually and services are provided on a cost recovery basis.

Consideration will also be given to developing a public use microdata file.

As described in Section 2, Statistics Canada, subject to provisions under the *Statistics Act*, will be sharing data on institutions and programs with a number of other organizations. Consideration is also being given to the development of a publicly accessible database containing this information.

Further documentation about PSIS is available at:

http://www.statcan.ca/english/concepts/PSIS/index.htm
http://www.statcan.ca/francais/concepts/PSIS/index_f.htm

APPENDIX
DRAFT TEXT FOR INCLUSION ON REGISTRATION FORMS AND IN CALENDARS

REGISTRATION FORMS

Under the federal *Privacy Act*, individuals can request access to their own, individual information held on federal information banks, including those held by Statistics Canada.

Students who do not wish to have their information used are able to ask Statistics Canada to remove their identifying information from the national database.

CALENDARS

Notification of Disclosure of Personal Information to Statistics Canada

Statistics Canada is the national statistical agency. As such, Statistics Canada carries out hundreds of surveys each year on a wide range of matters, including education.

It is essential to be able to follow students across time and institutions to understand, for example, the factors affecting enrolment demand at postsecondary institutions. The increased emphasis on accountability for public investment means that it is also important to understand 'outcomes'. In order to carry out such studies, Statistics Canada asks all colleges and universities to provide data on students and graduates. Institutions collect and provide to Statistics Canada student identification information (student's name, student ID number, Social Insurance Number), student contact information (address and telephone number), student demographic characteristics, enrolment information, previous education, and labour force activity.

The Federal *Statistics Act* provides the legal authority for Statistics Canada to obtain access to personal information held by educational institutions. The information may be used only for statistical purposes, and the confidentiality provisions of the *Statistics Act* prevent the information from being released in any way that would identify a student.

Students who do not wish to have their information used are able to ask Statistics Canada to remove their identifying information from the national database.

Further information on the use of this information can be obtained from Statistics' Canada's web site: http://www.statcan.ca/english/concepts/PSIS/index.htm, by writing to the Postsecondary Section, Centre for Education Statistics, 17th Floor, R.H. Coats Building, Tunney's Pasture, Ottawa, K1A 0T6 or by calling 1-613-951-1666.