

Maintaining Partnerships With Other Canadian Government Departments

By
Daniela Ravindra
Statistics Canada

**International Seminar on the Statistical Use of Administrative Data,
Seoul, May 19th – May 20th, 2008**

1. Introduction

Statistics Canada, the government agency responsible for collecting, compiling, analyzing and disseminating statistical information in Canada, has a long history of using administrative data in its programs. There has been a steady evolution over the years in the type of data used and in the ways in which these data are utilized. This paper will look at how Statistics Canada makes use of administrative data currently, the type of partnerships it has formed to be able to obtain these data and the legislative framework that allows it to obtain and use administrative data.

Although a large variety of administrative data files are used by Statistics Canada to produce estimates, this paper will focus on the use of business tax data as defined to include the income statements filed by firms for taxation purposes, as well as information from the payroll deduction system. Business tax data are received by STC from Canada Revenue Agency, the federal government department responsible for administering tax laws and various social and economic benefit and incentive programs delivered through the tax system for the Government of Canada.

2. History

Almost 30 years ago STC began investigating the potential for using business tax files for statistical purposes. In those days which preceded the advent of the computer, most information was received in paper form. This made for a more cumbersome process for using it: the data either needed to be captured by STC in the few existing automated systems, or the paper forms needed to be passed around to the analysts using the data. In consequence, the use of the information was minimal and limited to confrontation of survey data. This experience though allowed analysts at STC to discover the potential of business tax files as great sources of statistical information.

Later, as technology advanced, it became infinitely easier to obtain and process business tax data in the form of electronic files. Currently STC receives almost all the administrative data in electronic format. This means that the data are easy to load into STC databases, easy to process and to manipulate. It has also improved immensely the timeliness for the availability of business tax files. In some cases the data are available for analysts to use within hours of having been received by the Agency. Multiple users are also able to tap into the same database simultaneously, something which was not possible in the past. As a result, there was a proliferation of uses for business tax data: programs utilized it for imputing for missing survey data, editing of survey responses or estimating for non-surveyed portions of the business universe.

However, there wasn't a co-ordinated approach at the Agency level for all these uses. Rather each program acted more or less independently. This led to a lot of duplication of work and a lack of coherence between the ways in which various areas used the business tax data. To deal with these issues, in 1997, STC set up a separate group called the Tax Data Division. The initial role of Tax Data Division was to be the focal point for all work related to acquiring, processing and storing administrative data. It also became the liaison between STC and the Canada Revenue Agency, the department supplying business tax files.

A few years later, in the early 2000s, STC put forth a vision document which outlined the Agency's plan for expanding the use of tax data. This plan focused on four initiatives:

- expand use of tax data as replacement for survey records
- expand the use of tax data for edit and imputation
- fill data gaps (estimates for sectors of the economy not covered by surveys)
- demonstrate potential of tax data through analysis

The two main objectives of this plan were to:

- reduce response burden
- reduce STC expenditure on survey work

With this broader vision for STC, the role of Tax Data Division was also expanded: Most analytical projects assessing the quality of administrative data are now carried out under the auspices of Tax Data Division, as are the goals of further developing, promoting and facilitating the use of tax data at STC.

3. Establishing Partnerships

None of the work that is done currently at STC with business tax information would be possible without a close co-operation between STC and Canada Revenue Agency which has been the supplier of business tax data for the last 30 years. Early on, it was recognized by the Government of Canada's most senior bureaucrats that partnerships between Statistics Canada and Canada Revenue Agency would lead to benefits for the Canadian public. It would decrease the burden of responding to surveys, and would avoid duplication and expense associated with collecting the same information twice. Initial discussions at the deputy minister level led subsequently to agreements in principle for sharing administrative data.

The Statistical Agency took a gradual approach to establishing a partnership with Canada Revenue Agency. At first STC was involved in assessing the feasibility of using administrative data and so the relationship was of a more superficial nature. As the studies and the initial practical use of administrative data proved that this is a reliable and rich source of information, the two departments strengthened their relationships. Recently, as the Government of Canada has established ambitious targets for paper burden reduction on its citizens, there is an added incentive for co-operation¹.

In the early years, the various groups at STC who used business tax data dealt separately with their contacts at Canada Revenue Agency. As the use of tax data proliferated at STC the Tax Data Division became the focal point for all communication and liaison with Canada Revenue Agency. When a division needs to contact the Canada Revenue Agency, be it to obtain new data, to explore the potential of some tax form or to obtain any clarifications, they will go through Tax Data Division. This centralized approach allows for a more coordinated and efficient communication and delivery system. The two departments communicate on a weekly basis through the liaison groups to ensure that both of their requirements are met and to manage planned and unforeseen changes. Similarly, Canada Revenue Agency has positioned on their side the Statistics Division as the point of entry to their department.

4. Legal Framework Governing the Partnership

4.1 Legislation and Agreements

The partnership is governed by a formal agreement, a Memorandum of Understanding (MOU) signed by both parties. The MOU is based on the legal instrument that gives Statistics Canada its authority to collect information, The Statistics Act, and The Income Tax Act which outlines the mandate of the Canada Revenue Agency and its ability to share data with STC.

In particular, the Statistics Act states that the Chief Statistician, or his designate, may inspect and have access to any returns, certificates, statements, documents or other records obtained on

¹ Thirteen federal departments and agencies are participating in the Paperwork Burden Reduction Initiative (PERI). The objectives of the PBRI are to eliminate duplication, ease the reporting burden on businesses and demonstrate accountability.

behalf of the Minister of National Revenue as long as this information is subject to the same secrecy requirements to which it was subject when collected.

The MOU between STC and Canada Revenue Agency outline the roles and responsibilities of the partners. Specifically it includes:

- legislation under which the sharing of information can take place
- type of information shared
- the purpose for which the information will be shared
- who has access to the information
- the level of protection required for the information
- the sharing of the information with a third party
- dispute resolution mechanisms
- amendment process

4.2 Security of Information

The overriding principles of this MOU are that Statistics Canada will use these data exclusively for statistical purposes and that the confidentiality of the data will be protected. The confidentiality requirements are the cornerstone of the relationship between STC and any partner departments. It has always been emphasized at STC that protecting the confidentiality of data providers, be they respondents to surveys or tax filers is the only way to ensure the public trust and co-operation without which a statistical agency cannot exist.

As business tax micro data information is considered highly sensitive, added measures of protection are employed. Access to tax information at STC is granted on a “need to use” basis only and it is closely monitored. As well, the physical and electronic storage facilities for tax data are of the highest standards thus preventing disclosure or access by unauthorized employees.

Statistics Canada conducts regular audits on its internal procedures for safeguarding the administrative data and for sharing the data with third party users in order to ensure that all conditions in the MOU are fulfilled.

At no time does Statistics Canada send any information it has surmised from files back to the Canada Revenue Agency: the agreements ensure a one way flow of data only. This is another safeguard that reassures the owners of the micro-data (citizens and businesses) that the use of their personal and business information by Statistics Canada will not have any negative effects on their social benefits, tax paid or any other government program in which they participate.

4.3 Full Disclosure

In the interest of transparency, whenever possible, private citizens and business owners are informed of the fact that their personal and business tax information may be shared with Statistics Canada for statistical purposes only. This is done either by Canada Revenue Agency which indicates on their web site that they will be sharing the data with Statistics Canada or, in the case of taxation information, by Statistics Canada indicating on questionnaires and on user guides that taxation data will also be used for statistical purposes. Based on consultation over the years, STC has learned that the vast majority of Canadians support the use of administrative information by Statistics Canada. This is mostly based on two factors:

- trust in STC’s ability to protect the confidentiality of the data
- decrease in response burden

As long as these continue to be true, administrative data will continue to be an available source of information for STC.

4.4 Sharing Information with Third Parties

As head of the national statistics system, STC has a role to support the statistical activities of Canada's provincial governments. By sharing data with them it reduces response burden and contributes to increases in data quality.

Thus, STC forms partnerships with provincial statistical agencies and shares THEIR data with them, after obtaining it from Canada Revenue Agency. Only micro tax records for firms operating solely in that province are shared.

This process of data sharing has been enabled by amending the Income Tax Act and the Excise Tax Act. Statistics Canada controls the type of business tax data being shared, as it is responsible for safeguarding it once received from Canada Revenue Agency. To obtain administrative data from STC, provincial statistical agencies have to enter into agreements with STC. An MOU² is signed with each province before data can be shared.

5. Maintaining Partnerships

Once a partnership has been established and an MOU is in place, it falls on the two partners to maintain this relationship. Some partnerships between Statistics Canada and other provincial or federal government department are mutually beneficial; others only meet a Statistics Canada need. In the case of mutually beneficial partnerships, the department supplying the information has a vested interest in the sense that it will ultimately use the information published by Statistics Canada for analysis and policy making. However, in the case of Canada Revenue Agency the department sharing the information does not derive any direct benefit from having shared information with Statistics Canada. This is why it is crucial for this type of partnership that there be support from the highest level of management at the Canada Revenue Agency. This level of support must be transmitted down the line to the working-level staff so that they understand the importance of this work.

To ensure that the partnership is well maintained a structure of governance has been established. Senior managers at the assistant deputy minister level from both departments meet on a regular basis twice a year to discuss the broad issues affecting the partnership, while working-level staff maintains very close communication on a weekly basis to ensure that operational issues are dealt with and to plan for future activities. As well, directors of stakeholder divisions on both sides meet a few times a year to keep the communication flowing.

It is important that both partners are equally involved in maintaining the relationship and that, in times of crisis, one partner is ready to accommodate the other. STC has prepared comprehensive contingency plans for all its outputs which rely heavily on administrative information. These contingency plans are shared with Canada Revenue Agency.

There are a number of ways through which STC can deal with unexpected problems related to the delivery of files by Canada Revenue Agency:

- adjusting the delivery of its own products when possible
- lending the help of STC staff
- adjusting the content of its own products when possible
- using other sources of information
- using historical data, etc.

² These MOUs contain the same stringent confidentiality protection requirements as the one signed between STC and Canada Revenue Agency.

Both departments keep each other informed of any planned changes. Regular meetings at the working group level are essential in these cases. In certain cases, when changes are planned, Statistics Canada is invited to review the plans and contribute an opinion. There have been instances when redesigns of forms used to administer programs at Canada Revenue Agency have included Statistics Canada requirements. This is done in recognition of the fact that, as part of the Government of Canada, all departments have the obligation to work together to ensure that public resources are used in the most efficient way for the benefit of all Canadians.

In the last few years STC has obtained from Canada Revenue Agency the approval to call business tax filers to confirm information received. This is done now routinely for STC's Labour Statistics program which uses payroll deduction information. Most recently internal discussions at STC have suggested that this type of procedure could be extended to other business programs.

When issues cannot be resolved at the working level, they are escalated to management and then to senior management if necessary.

In summary, establishing and maintaining partnerships between Statistics Canada and other departments is a complex process which requires dedication and perseverance on both sides, and a deep belief in the public good

6. Type of Business Tax Data Used

The tax and employment insurance data files used by STC can be grouped into two general categories:

- principal files used by a very large number of programs in a more or less similar fashion
- files that are specific to the use of a more specialized program

The main files are:

- tax file containing income tax statements of incorporated business
- tax file containing personal income tax information as well as business income information of owners of unincorporated businesses
- payroll file containing wages and salaries paid to individuals, as well as employment insurance deducted
- summary of payroll information containing total of wages and salaries paid by a firm as well as number of employees
- the Goods and Services Tax (GST) which is a value added tax file

The information contained in these files is used to produce industry estimates of revenues and expenses, estimates of wages, salaries and number of employees and to provide universe data, as mentioned before for the System of National Accounts. The personal income tax file is also used to derive inter-provincial migration and income and demographic statistics for individuals and families.

All these files have been in use for a considerable number of years at STC. Tax Data Division has developed over the years a very robust system of validation edits and imputation that is applied to these files before they are used by specific subject-matter programs. Numerous studies and comparisons to survey data have been done in the past to validate the accuracy of these tax files. They are considered to be presently as reliable as data collected through survey.

The only exception to the above would be the GST file which has been in use for a relatively short period of time and for which the edit and imputation system is still being improved. More research needs to be done on this file to determine how to correct for some of the issues it presents (infrequent reporters, large revisions, etc.).

There are literally hundreds of other business tax files that STC receives. Among those, many are used by a few areas for very specific purposes. Some examples include:

- the Child Tax Benefit file
- file containing research and development expenses for businesses
- pension plan files
- files of transactions with non-resident partners
- files containing information on non-profit organizations

Quite often these files are edited and validated by the analysts responsible for the program which will be using them. They are not subject to a generic system of validation edits. The plan is that as the files become used by a larger number of programs, a more centralized approach to processing them would be used. A key part of STC's quality assurance practices is to ensure that the tax files used are coherent with survey data and that the methodology employed to process these files is standardized. This not only eliminates inconsistencies across programs, but it also avoids duplication of work.

Currently 34 divisions which conduct over 200 business surveys at STC use tax files in their processes.

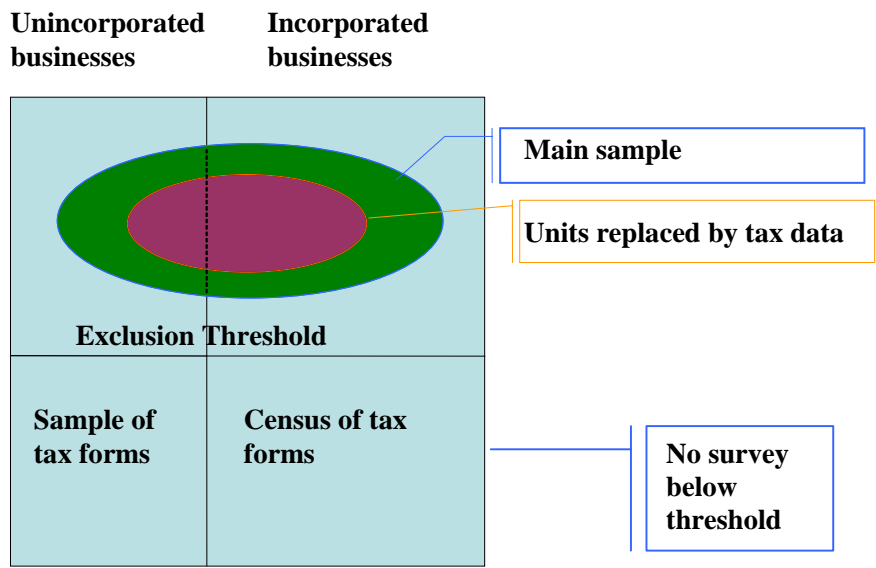
7. How Does STC Use Business Tax Data

As mentioned before, the use of business tax data at STC has gone through a long evolutionary process. Currently, nearly all business programs rely heavily on tax data. They include manufacturing, retail trade, services industries, agriculture and international trade.

7.1 Replacement of Survey Records

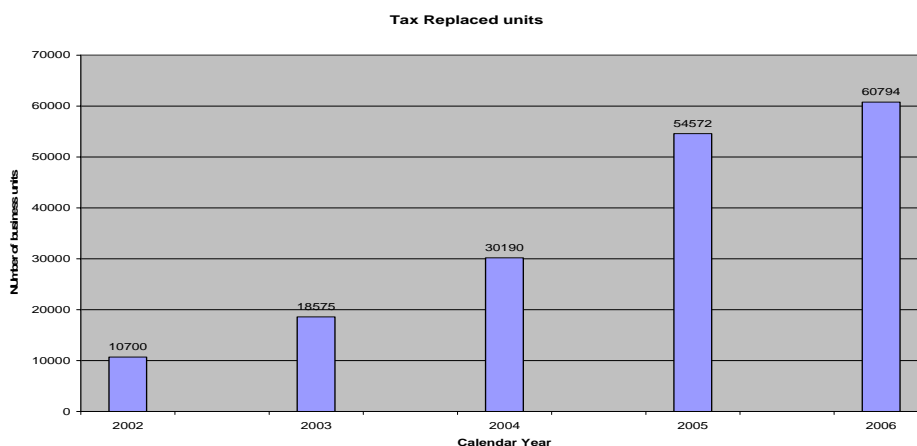
The most common use of business tax data is that of replacing surveyed units. There are two methods for this: simple direct replacement used for annual surveys and replacement aided by a methodological model for sub-annual surveys.

The diagram below depicts the use of tax data for most annual business surveys. The universe is made up of incorporated and unincorporated businesses. Before any tax replacement occurs, the sample has to be selected. Normally, very small units, accounting for less than 10% of the total revenue (exclusion threshold) of each stratum are not considered for sampling. Tax information is used to estimate for this portion of the universe. For the incorporated units STC receives all tax forms therefore a census of these forms can be used. For the unincorporated businesses, STC obtains only the electronically filed tax forms and therefore an estimate is calculated from those.



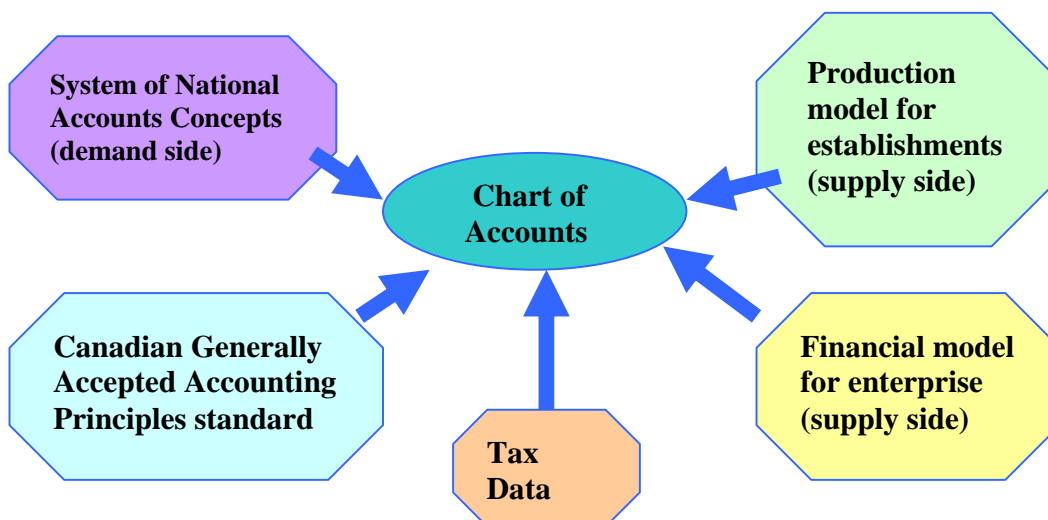
Once the sample is selected from the remaining units, a certain portion of them are tagged to be replaced by tax records. In general, the majority of the units replaced by tax are small or medium in terms of revenue size and their structure is simple: only one establishment in operation. Business units which have more than one establishment are more difficult to replace by tax, since fiscal returns are normally filed for the entire business operation. In cases when business activities cover more than one industry, the business tax data would have to be allocated to each industry and this is not always a straightforward activity.

As all of STC's business surveys are mandatory, the use of business tax files for survey replacement contributes a great deal to decrease in response burden. Currently, over 60% of simple units in sample for business surveys are replaced at STC by tax information. The table below shows that between 2002 and 2006 the number of tax replaced units has risen six fold to 60,794. They represented close to 10% of the total gross sample size for 2006 and contributed to a significant decrease in the overall response burden on businesses.



7.1.1 The Chart of Accounts

There is not always an obvious correspondence between the survey variable (asked on the questionnaire) and the variables reported by companies on their income tax returns or their employment insurance forms. There can be differences in concepts and definitions since surveys are generally measuring production, while the tax files contain revenue and expense information reflecting the operations of the company. In order to ensure that tax data are used in a consistent and correct manner across all business surveys, and that tax as well as survey data are correctly linked to the production concepts and to the System of National Accounts concepts, STC developed a standard way of relating these various sources by mapping them to a standardized income statement. This standardized income statement is known as the STC Chart of Accounts (CoA). The CoA is compliant with the Canadian Generally Accepted Accounting Principles.



This exercise has also led to, in some cases, the adaptation of some survey concepts to better suit the needs of the businesses. It was discovered that asking respondents for information similar to what they report on their income tax led to a better response rate. An additional benefit was a greater consistency across business survey questionnaires.

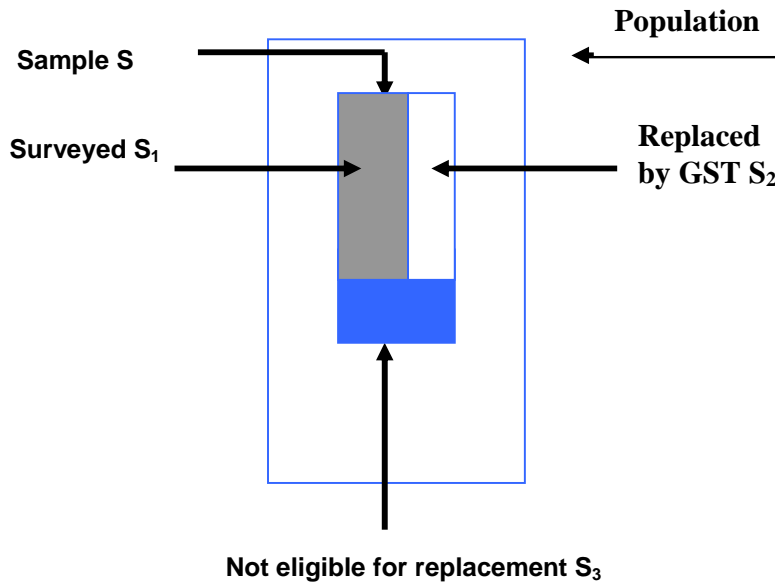
Not all business surveys at STC are 100% CoA compliant, but efforts are being made to accomplish this in the near future.

7.1.2 The Methodological Model

For sub-annual estimates, STC relies on information coming from the value added tax (GST) files. These are received on a monthly basis and have been used for the last 5 years to replace survey data in the monthly manufacturing, retail and services surveys. However, just as the annual surveys required the CoA to make the link between the tax and the survey data, for sub-annual estimates a mathematical model is necessary to bridge the two. There are two aspects to this model:

- one that adjusts for the difference in concepts/definitions
- one that adjusts for the timeliness

The timeliness of the GST files is not good enough to meet STC deadlines, so a previous month's data has to be used. Hence, the necessity for the mathematical model to adjust for the time lag.



The diagram above shows how the sample for a monthly survey has been divided into three portions: a portion eligible for tax replacement but which will be surveyed (S_1), a portion to be replaced by GST records (S_2) and a third portion which is made up of complex units which will never be available for tax replacement and therefore not used in the model.

The model is based on the relationship between the revenue obtained from survey and that from the GST for the units that were available for replacement but remained in the surveyed portion (S_1). The adjustment derived from this relationship is then applied to the GST revenue values for those units that were selected for replacement (S_2). The trend for the surveyed portion S_1 between m and $m-1$, where m is the month of reference, is used to adjust for the time lag.

$$S_{2, m-1} = \frac{GST_{ns, m-1} \times S_{1, m-1}}{GST_{s, m-1}}$$

Once the S_2 at $m-1$ is obtained, it must be adjusted to obtain the current period value. This is done by using the trend for the surveyed portion S_1 :

$$S_{2, m} = S_{2, m-1} \times \frac{S_{1, m}}{S_{1, m-1}}$$

This model adjusts for both the conceptual differences between GST and survey revenue, as well as for the timeliness issue. It is being applied currently for three STC sub-annual surveys: manufacturing, retail trade and food services with very good results.

7.2 Data Processing and Data Confrontation

When STC first started using administrative information, it was for confronting survey data. It is still the most widespread use of administrative data at the Agency. Every business program, as part of its data validation exercise, will look at tax data either at the micro or at the macro level to validate survey results. Most data analysis tools at STC contain tax and survey information side by side to help analysts in their work.

Editing respondent data and filling in for non-response are also two processes that employ tax data before most other information is used. In fact, for many business surveys the database is first populated with tax information for all the variables which are available and then, as survey information starts coming in, the tax data are overwritten. Most business surveys at STC are processed through a fairly sophisticated edit and imputation system that applies deterministic, donor and prorating functions which will use tax data as inputs. In most of these cases tax data have been reported, but edited data are also widely used.

7.3 Data Gaps

Getting additional information without increasing response burden is one of the most daunting challenges faced by STC. The demand for information is increasing continually, while the willingness of business owners and private persons to respond to surveys is diminishing. Once again, administrative data provided a possible solution to this dilemma. Currently, a number of programs at STC are entirely tax based (e.g. some agriculture programs), while others alternate between tax and survey every so many years (e.g. Survey of Taxes and Limousines).

An important project is now under way to develop a methodology for producing quarterly estimates based entirely on the GST file. Tests have shown so far that for many industries this would be a feasible solution, but not for all due to a certain lack of timeliness in receiving the tax information. These estimates will most probably be used as trend indicators, while more robust annual surveys will provide the benchmark levels.

7.4 Universe Data

The System of National Accounts relies on tax files to provide them with an estimate of the complete universe of businesses and individual tax filers. Not all industries in Canada are covered by a survey program, and also for the industries covered often the small contributors are missing. So, in order to be able to produce a national figure for the Gross Domestic Product, for example, tax and employment insurance files are used as proxies for the missing portions. Wages and salaries information are of particular interest when it comes to estimating growth for certain industries.

7.5 Business Tax Data Related Training

It is extremely important that users of business tax data have a good understanding of the tax file content. This includes the tax concepts as well as all the processing steps through which the raw

tax files received from Canada Revenue Agency are transformed into a coherent, complete and accurate set of files. To this end, STC has invested a considerable amount of resources in training users on all matters related to tax data. A number of courses are offered to staff on a regular basis, and ad-hoc, more specialized training is designed for specific users. Additionally, a comprehensive set of metadata has been developed over the years for variables present on the files which are more widely used. These efforts are key to ensure that business tax data are used correctly and that staff also have the opportunity to learn about tax information they could potentially find useful but they do not currently use.

8. Conclusion

Even though the use of business tax data at STC is well established, the Agency is still pursuing research into other innovative uses. It is now at the stage where tax variables or files not currently in use are investigated for very specialized uses, and where files currently in uses are being tested to see if they are robust enough for producing trend estimates that are 100% tax based.

There are also a number of proponents of a tax based business program who would like to see a statistical system based on tax data first, where survey information would only fill in for information not found on tax files. In short, STC is constantly pushing the envelope on the use of business tax data because what has been accomplished so far is seen as only a fraction of what it is possible to accomplish.

References

1. Brodeur, M. and D. Ravindra. 2007. Unified Enterprise Survey-New Horizons, Statistics Canada.
2. McMechan, J. 2008. The Use of Tax Data in Business Surveys, Statistics Canada Course
3. Ravindra, D. 2008. SIMP II Economic Statistics Document Series - Document # 22: Building Partnerships with Other Departments
4. Statistics Canada 2007. Statistics Canada's Paperwork Burden Reduction Initiative Report
5. Statistics Canada 2008: Statistics Canada's 2006 Estimates of Compliance Costs and Response Burden Reduction Initiatives
6. Statistics Canada 2007: Tax Data Division Quadrennial Program Review