THE USE OF ADMINISTRATIVE DATA FOR INITIAL AND SUBSEQUENT PROFILES OF ECONOMIC ENTITIES

Colleen Clark and Robert Lussier, Statistics Canada

Statistics Canada is currently rebuilding its central register of economic entities. The new register views each economic entity as a network of legal and operating entities which define statistical entities. This network view, the profile, is determined through the 'profiling' process which involves contact with the economic entity. In 1986 a list of all entities in-scope for a profiling contact was required so that profiles could be obtained to initialize the new register. Administrative data were used to build this list. In the future, administrative data will be a source of information on changes that may have happened to economic entities. They may thus be used to request review and updating of profiles.

The paper begins with the objectives of the profiling process. The procedures for constructing the frame for the initial profiling process using several administrative data sources are then presented. These procedures include the application of concepts, the detection of overlap between sources, and the evaluation of data quality. Next, the role of administrative data in providing information on changes to business entities and in requesting profiles to be verified is presented. Then the results of a simulation study done to assess this role are reviewed. Finally, the paper concludes with a series of questions on the methodology of using administrative data to maintain profiles.

1.0 INTRODUCTION

Statistics Canada is in the process of reorganizing its programme of economic surveys. The new programme will result in an increased use of administrative data. They will be part of a Central Frame Data Base (CFDB) from which economic surveys will draw their sample. They will also be used to maintain the CFDB. This and other elements of the reorganization strategy are contained in Colledge and Lussier (1985). Experiences in the implementation of the strategy are contained in Colledge (1987).

One of the first steps in the reorganization was to formulate definitions of the CFDB units. A fundamental unit is the business entity. A business entity is defined in Statistics Canada (1987) as 'an economic transactor having the responsibility and authority to allocate resources in the production of goods and/or services, thereby directing and managing the receipt and disposition of income, the accumulation of property, the borrowing and lending of capital, and the maintenance of complete financial statements accounting for their responsibilities.'

The reader should note that the term 'business,'in the United States is often restricted to the trade sector. This is not the case in Canada. Business entities cover all sectors of the economy, including areas such as manufacturing, transportation, and professionals. The term 'economic' was used in the title to avoid a potential misunderstanding. However, the term 'business' is used throughout the paper.

The Central Frame Data Base currently being built by Statistics Canada attempts to represent the structure of the Canadian economy. It recognizes that this economy is dominated by a small number of large business entities who account for the majority of the activity within the economy. The CFDB is divided into two components paralleling this dichotomy.

One component, the Integrated Portion (IP), provides coverage of the small number of large or otherwise important business entities, while the other, the Non-Integrated Portion (NIP), covers the remaining large number of smaller entities. The entities in the former component are more complex. Hence, the identification of those portions of the complex business entity that are of interest to a particular survey requires substantial effort.

The Integrated Portion (IP) of the CFDB attempts to represent the complex structure of business entities through the use of an Information Model. The model consists of five structures linked together which describe a business entity. These structures allow survey populations to be accurately identified. Entities on three of the structures are not controlled by Statistics Canada,while entities on the other two are generated by Statistics Canada for the purpose of collecting, editing, estimating, and tabulating economic data. The five structures are:

- The legal structure , which describes the legal representation of the business entity. It is comprised of legal entities and their relationships of ownership and control. Examples of legal entities are incorporations under federal or provincial charter.

- The operating structure, which describes how the business entity operates and how it organizes its accounting system. It is comprised of operating entities. This structure organizes and controls the production of goods and/or services. It is an attempt to structure the business entity as it sees itself. Examples of operating entities are divisions, profit centres, and plants.

- The statistical structure , which consists of a hierarchy of statistical entities. These entities are derived from the associated operating structure, depending on the units within the operating structure for which records for a particular set of data are maintained.

- The reporting structure, which consists of reporting arrangements for each selected statistical entity by survey. The data available in the accounting system of the business entity are collected from the reporting entities.

- The administrative structure, which contains administrative data such as income tax data collected from legal entities and payroll deduction account data collected from operating entities.

The complex process of determining the boundaries of the business entity and of delineating its five IP structures and their associated links is termed 'profiling.' This network view of the business entity is the 'profile.' The data to construct a profile are obtained through a contact with the business entity or some component of it. The entity's legal and operating structures,as well as some administrative structure data items, are obtained, or reviewed and updated, during the interview. The statistical structure is then generated or updated automatically from the new operating structure. Finally, default reporting entities are created for new selected statistical entities using selected fields from the legal, operating or administrative structures. These entities may subsequently be updated as a result of the first survey

contact with the respondents.

The type of profiling contact used depends on the entity's complexity and any special reporting arrangements. The most complex and important entities will receive a personal visit from either Head Office or Regional Office personnel. The remaining entities will be contacted by telephone. Entities will be contacted about once every two years, or more often, depending on how quickly their structures change.

Cyclical profiling, whereby business entities are periodically contacted, is one method that will be used to keep the IP of the CFDB current. A survey feedback process and data from administrative sources will also be used.

The design and construction of the CFDB is taking place over three years, culminating in a data base that will be available for integration into survey programs in April 1988. Most of the data in the Integrated Portion of the CFDB in April 1988 will have come from a profiling process that began in April 1986. However, no single list of business entities in-scope for a profile was available in April 1986.

Administrative data played a major role in initiating the profiling process. It was used as a starting point to construct the current Statistics Canada view of the business entity. A list of business entities in-scope for an initial profile was assembled from administrative data sources. Section 2 describes how this was accomplished. Section 2.1 gives the frame requirements. A description of the data sources used to build the frame follows in Section 2.2. Section 2.3 shows how the frame unit was constructed and how the various data sources were combined to build the frame.

2.0 USE OF ADMINISTRATIVE DATA FOR INITIAL PROFILING

2.1 Frame Requirements

The first step in building the frame for initial profiling was to define the frame unit. The ideal one would be the business entity. However, this entity was not available either internally or externally to Statistics Canada. The units available to us were essentially legal entities. It was necessary, then, to group legal entities to approximate business entities. The frame unit was defined as a grouping of legal entities subject to the following constraints:

- The definition of the business entity implies that it covers all legal entities linked through control. One type of control is established through owning more than 50% of the voting rights of a legal entity. The grouping of legal entities through this control rule is restricted to one level of foreign control outside Canada.
- There has to be a single Canadian legal entity that controls all other Canadian legal entities in the business entity. This is necessary because profiling contacts with the business entity could only be made in Canada.

The next step was to determine which frame units would comprise the frame and what data were required for each. The frame from which business entities would be selected for an initial profiling contact and from which the initial picture of the business entity would be generated would contain all business entities in scope for a contact.

Business entities are in scope for a profiling contact if they qualify to be members of the Integrated Portion

of the CFDB. Membership is determined by criteria applied to the legal structure that describes the legal representation of the business entity.

Legal structures can become members of the Integrated Portion in one of two ways. First, if the structure consists of only one legal entity, then the legal entity is part of the Integrated Portion if its revenue during its fiscal year of interest is above a set prespecified value. This prespecified value depends on the legal entity's major industry and the location of its head office. Alternatively, if the legal structure consists of more than one legal entity, then the legal structure is part of the Integrated Portion if at least one of the legal entities in the structure is above its appropriate prespecified value.

Therefore, in order to determine which business entities are in scope, the following information was required for every legal entity:

- Relationships of ownership between legal entities.
- Revenue in the fiscal year of interest, primary industry, and head office location.

For business entities that qualify to be on the frame and, hence, to receive an initial profiling contact, information was required to select and contact the entity. The following was required to select the entity:

- All industries in which the business entity was involved so that the Wholesale and/or Retail industries could be contacted first. The surveys of these industries required a set of statistical entities that had been generated from a profiling contact before other surveys did.
- The number of physical locations of all business entities that consist of one legal entity or that consist of two legal entities of which the owner is foreign. This data item determined the type of profiling contact that would be made as either a telephone contact by Regional Office staff or a personal visit by Regional or Head Office staff.
- The province in which the ultimate Canadian owner was located. The province was used to distribute the workload of making the profiling contacts to regional offices according to their capacities.

In order to contact the business entities, name and address were required for the legal entity at the top (excluding foreign owners) of the business entity. Contact data and any special reporting arrangements that surveys had recently used would be desirable.

2.2 Data Sources

The data sources which could be used were restricted, primarily, by the frame coverage requirements. This restriction eliminated sample lists and many industry-specific lists such as survey frames. Only data sources that were lists of all legal entities potentially in scope for a profiling contact that carried, at least, some of the required data items could be considered. These data sources were:

- The Inter-Corporate Ownership Database (ICO), which is a list of all legal entities operating in Canada that are owned by either foreign or Canadian legal entities and their owners. The coverage of foreign legal entities is to the extent required to determine the ultimate owner.
- The Current Business Register (BR), which is primarily a list of all legal entities that are employers. The number of physical locations of a legal entity, contact data (address and reporting

arrangements) used by surveys, and the industries in which the legal entity operates are available here.

- The Corporation Tax Base (CORP) which is a list of all legal entities that filed a corporate tax return with Revenue Canada, Taxation in a given year. The primary industry, the location of the Head Office, and revenue for the fiscal year are carried on this data source.
- The Individual Tax Base (IND) which is a list of all individuals who filed a tax return with Revenue Canada, Taxation in a given year. Individuals who report self-employed income on their return are legal entities of interest to Statistics Canada economic surveys. Primary industry data and contact data are available from this tax base for each individual reporting self-employed revenue as is his/her revenue from self-employment.

Both of the tax base data sources (CORP and IND) are administrative data files. Administrative data received monthly from Revenue Canada, Taxation regarding an employer's payroll deductions are used to update the BR. The ICO data source is a census survey response file.

None of these data sources provides complete coverage and all the required data items. Rather, coverage can only be obtained by combining these data sources. The same is true for some required data items, while for the rest more than one source can provide them. The strategy used to combine these data sources to obtain the best coverage and data quality is presented in the next section.

2.3 Frame Creation Procedures

The challenge in creating the frame for initial profiling contacts lay in integrating four data sources that had each been designed for different purposes and had never been integrated to this extent before. This situation is common to users of administrative data. The task was even more complex because this was the first time many concepts established for the CFDB were applied.

The constraints of limited time and resources forced the project team to make some assumptions when creating the frame. However, the assumptions were justifiable since the picture used on the frame would be corrected through the profiling process. A simple description of the procedures used is presented in this section.

There were three steps in the frame creation process, each of which is discussed in the following sections.

- Construct a list of all potential frame units;
- determine which are in scope; and
- acquire selection and contact data.

2.3.1 Create Potential Frame Units

The frame unit was constructed by grouping legal entities in the following manner to create business entities. The legal entities were first grouped into legal structures. One legal structure consisted of that set of legal entities related via ownership of more than 50%. Relationships involving foreign legal entities were accepted only if the foreign legal entity owned or was owned by a Canadian legal entity. When a foreign entity owned more than one Canadian entity, the legal structure was divided into as many business entities as there were Canadian entities directly owned by the foreign entity. In this way, a profiling contact would be made with the ultimate Canadian owner of each resulting business entity. Examples are provided in

Diagram 1.

Individuals who reported self-employed income were considered as a legal structure containing only one legal entity. The ownership of corporations by individuals, as well as relationships of joint venture between corporations, were not considered in constructing business entities.

DIAGRAM 1

DEFINING BUSINESS ENTITIES

EXAMPLE 1
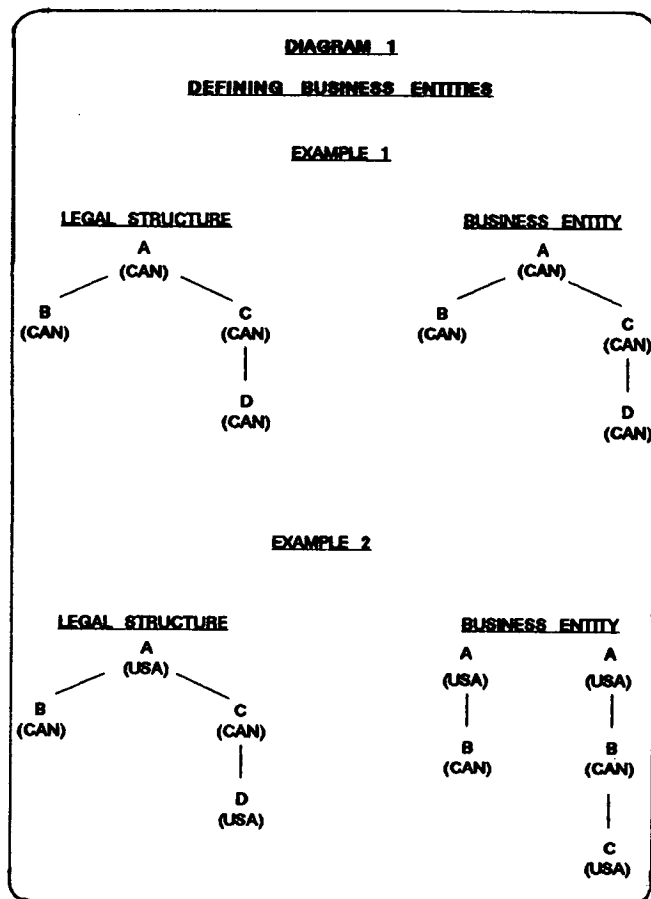
Therefore, we can think of the set of business entities in scope for an initial profiling contact as two mutually exclusive groups. The first group consists of legal entities that represent individuals who report self-employed income. The Individual (IND) tax base contains a list of all potential frame units in this group.

The second group consists of legal entities that represent corporations operating in Canada. The Inter-Corporate Ownership (ICO) data source was manipulated to provide a list of corporations that belonged to legal structures containing more than one legal entity. A list of all legal entities that are not owned by any other legal entity was obtained from the Corporation tax base, after elimination of those legal entities that were owned by other legal entities or were owners themselves. That is, it was necessary to match the ICO source and the CORP Tax base to identify the overlap between them. Legal entities that appeared on both sources could thus be identified to ensure that they would only appear once on the frame. Linkage between the two sources was not straight-forward and involved a clerical process because a common identification number was often not available.

2.3.2 Determine In-Scope Frame Units

The data required to determine if individuals reporting self-employed income were in scope was on

the IND tax base. It was a simple step to determine if a legal entity was above its appropriate prespecified cut-off.

The situation was more complex for corporations. The linkage achieved between ICO and CORP provided the data required to apply the cut-off rule. However, about 20% of the corporations on ICO could not be linked to CORP. In these cases an assumption was made which led to an overestimation of the set of business entities in scope for an initial profile. It was assumed that legal structures which contained at least one unlinked corporation satisfied the frame inclusion conditions. Otherwise, legal structures were frame members if at least one corporation satisfied the cut-off rule.

2.3.3 Acquire Selection and Contact Data

The result of the previous step was a proxy list of all business entities in scope for an initial profiling contact. The data required for selection and contact described in Section 2.1 that are not already on the frame were available from the BR. The frame and the BR overlap because a majority of the frame units representing corporations, and a smaller proportion of the frame units representing individuals, are employers. Linkage between the frame and the BR was required so that data from the BR could be added to the frame for units found on both sources. That is, it was necessary to detect duplication between the two sources.

It was even more difficult to link these two sources than it had been to link the ICO and CORP sources. This was due not only to the frequent absence of common identification numbers, as in the ICO-CORP case, but also because the BR resembles a business entity's operating structure more than its legal structure. The name and address from the BR were used for linking when no common identification number was available. However, the names and addresses on the BR often refer to 'trade' or 'operating' locations, which are sometimes different from the 'legal' names and addresses on the ICO and CORP sources. When this occurred it was difficult to establish a link and, hence, eliminate duplication.

There were some frame units for which no link to the BR was achieved, either because they were non-employers and therefore not on the BR or the linkage procedures could not establish the link. In these cases subsequent stages in the initial profiling process were amended to accommodate the frame limitations. Contact data of a lesser quality were taken from the tax base. The selection criteria were changed to reflect the absence of data on industrial breakdown and physical locations for these legal entities.

When a legal entity was involved in only one industry, the primary industry was available from both the tax bases and the BR. It was necessary, then, to reconcile this common data item when they were different. In this case the BR industry was used, since it was considered more reliable.
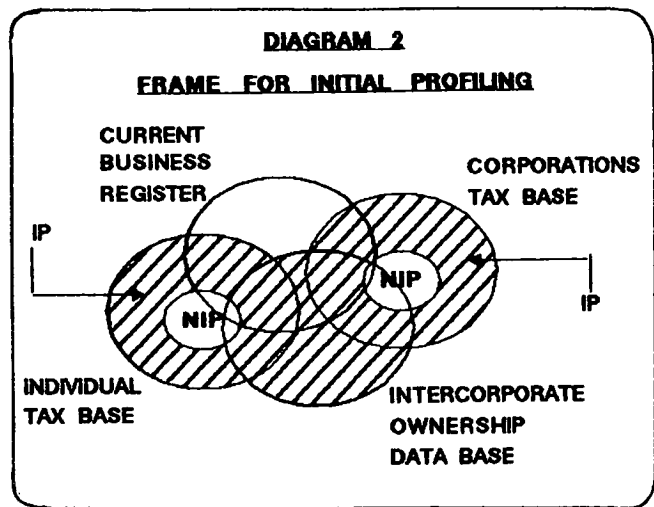
A pictorial representation (not to scale) of the resulting frame is shown in Diagram 2.

2.3.4 Evaluate the Frame

The quality of the resulting frame was assessed by three projects. First, the consistency of the frame with the specifications for creating it was verified.

The second project involved comparing various distributions of the legal entities on the frame with the same distributions produced from an independent simulation of the Integrated Portion. The distributions did not differ significantly.

Lastly the frame was assessed by comparing it with

DIAGRAM 2

FRAME FOR INITIAL PROFILING

CURRENT BUSINESS REGISTER

CORPORATIONS TAX BASE

IP

NIP

NIP

IP

INDIVIDUAL TAX BASE

INTERCORPORATE OWNERSHIP DATA BASE

the BR. A sample of 30 of the larger units in the BR was matched to the frame for initial profiling. All of the entities were found, but with great difficulty, because the two sources use different concepts.

2.4 Conclusion

The frame strategy just described was based on some simplistic assumptions regarding coverage, data quality, and the way in which business entities operate. 'Shortcuts' were often used to satisfy the frame requirements. It was felt that this approach was justified because of the role of the frame as a provider of initial pictures of business entities that would be updated during the profiling process. The implications of making these assumptions are discussed in this section.

The population of business entities in scope for an initial profiling contact may contain duplicates and out-of-scope units. If so, then more profiling contacts than necessary will be made. This would increase Statistics Canada's production costs. It would unduly burden the respondent with duplicate requests. Finally, the image of Statistics Canada could be adversely affected.

The population may be underestimated. Nevertheless, the missing units will be profiled at a later date. This would delay the introduction of new large units into the Integrated Portion of the CFDB. The missing units would be covered by the Non-Integrated Portion in the interim rather than the Integrated Portion.

Inaccurate selection and/or contact data could complicate or delay contact until accurate data could be found. The consequence in these cases is also an inaccurate CFDB until the profile is completed.

These experiences demonstrate the complications introduced when administrative data are used. They also illustrate the care that must be taken in ensuring the compatibility of administrative data with one's requirements. Examples were provided of the types of ensuing compromises that must be made when a reasonable compatibility cannot be reached.

3.0 USE OF ADMINISTRATIVE DATA IN SUBSEQUENT MAINTENANCE PROFILES

3.1 Cyclical and Reaction Profiling

There will be two types of subsequent maintenance profiling, namely cyclical and reaction profiling. Each of these is explained below.

Cyclical profiling is the process that will ensure

that all business entities in the profile population get reprofiled within a certain period of time. It is expected according to current budget forecasts that this period of time will be two years. Time elapsed since the business entity's last profile will be the factor that determines eligibility for cyclical profiling. Other factors will be taken into account to prioritize the eligible units within cyclical profiling.

Reaction profiling is the process that will profile a business entity as a result of information through a source other than profiling that changes may have occurred to that business entity and that the statistical image of the business entity on the register may not be valid any longer. Reaction profiling will keep the CFDB more up-to-date than if only the cyclical profiling mechanism were used. Some of the sources of information on changes are the various files of administrative data received regularly at Statistics Canada.

3.2 Sources of Administrative Data that Can Be Used

The three sources of administrative data that Statistics Canada can use to update its central register that are discussed in this paper are:

- the Individual Tax Base;
- the Corporation Tax Base; and
- data on payroll deduction accounts captured by the tax authorities.

Generally, individuals and corporations file a single tax return for a reference year. However, it is possible to have more than one return for a reference year if, for example, a corporation changed its fiscal year end with the approval of the tax authorities. Nevertheless, one can say that tax returns are an annual source of changes.

The receipt of the tax bases at Statistics Canada does not occur at a single point in time. In fact, Statistics Canada receives files of tax data regularly for a reference year over a period of two years. Thus, one could perform monthly updates to the register from tax data but each register record would generally be updated only once a year.

On the other hand, an employer is generally expected to send remittances for his payroll deduction accounts on a monthly basis. In turn, Statistics Canada receives a file of payroll deduction account data once a month. Thus, monthly updates can be made to the register from payroll deduction account data and each register record can in theory be modified every month.

Note that there are other sources of administrative data that could be used. They are not discussed in this paper because they are not obtained on a universe basis or on a regular basis. They are, nevertheless, worth mentioning. These are:

- limited information on corporations that have not filed a tax return but are believed to be active, captured by the tax authorities;
- additional data captured from a sample of tax returns by Statistics Canada; and
- data on a tax authority form filled out by employers when they request a payroll deduction account, captured by Statistics Canada.

3.3 Signals of Change

Signals of change were developed from the administrative sources described in the previous section. These signals identify administrative records for which changes to their associated statistical entities may have occurred. They also inform the register that reaction profiling may be desirable for these entities to keep the register up-to-date.

The signals are administrative source dependent. For each of the three sources listed in 3.2 the signals consist of comparison tests between new data received for an administrative record and the last data received for the same record from the same source. These tests may involve a single field or a group of fields and may be conditional on a single field or several fields. These comparison tests attempt to identify real world events that have an impact on the statistical entities and not only on the administrative entities. Remember that the statistical entities exist for the purpose of economic statistical programs and often are completely different from the legal-administrative reality. Therefore, these comparison tests should optimize the detection of changes in the administrative data that reflect a change in the statistical entities. As an example, change of ownership of a manufacturing plant may mean the 'deathing'of an administrative record and the birthing of a new one. On the statistical entities, it may however mean no change, as the same establishment with its capabilities to provide the required data may still exist.

If the frame was updated directly from the changes noted in the administrative records, the consequence would be a high incidence of apparent deaths and births in the statistical entities and a risk of incomplete or duplicated coverage. Thus, there is a requirement to contact respondents, or at least to perform in-house research using all available documentation, to find out for signaled administrative records what happened to the statistical entities. The "translation" process is not trivial at all and its resolution constitutes the purpose of reaction profiling.

The number of signals that were determined from each source, together with some signal examples, are presented in Table 1. One should however note the following points in studying the data on the number of signals. Some signals are very refined, while others are not. It was often decided to split an original signal into mutually exclusive sub-signals, because it was felt that it may be more informative in determining the action to take from the signal. The most trivial example concerns the Payroll Deduction Accounts. Eighteen of the 40 signals represent changes in the estimated number of employees covered by the account. The 18 signals distinguish between increases and decreases in the estimated number and the magnitude for each of them. It was thought that such a breakdown would be informative to prioritize the clerical work. Nevertheless, one could consider these signals as one.

TABLE 1
SIGNALS BY ADMINISTRATIVE SOURCE

Adminis-trative Source	Number Of Distinct Signals	Examples
Annual individual tax returns	50	Change from single province of taxation to multiple jurisdiction
Annual corporation tax returns	49	Start of a joint venture
Monthly payroll deduction accounts	38	New account with descriptions in the name that identify a corporation

It is expected that even though tax returns are processed regularly, a given return will generally generate signals at most once per reference year, while a given payroll deduction account may generate a signal or signals every month. What is of more interest, therefore, is not the number of signals defined per source but the number of records that are identified by these signals. This would give an idea of the amount of clerical resources that will have to be invested to update the register from administrative sources. A simulation study was thus undertaken to address this issue.

3.4 Simulation Study

The simulation study consisted of applying the signals previously described to the following populations:

● the individual tax returns for fiscal periods that ended in 1984, to detect changes that had taken place during these periods;
● the corporation tax returns for fiscal periods that ended in 1984, to detect changes that happened during these periods;
● the payroll deduction account of the beginning of October 1985, to detect changes that had occurred since the beginning of September 1985.

The results of the simulation study are presented in Table 2. The following observations can be made on the results:

● There are a very large number of tax returns that generate signals: only about one eighth of the individual tax returns and one fifth of the corporation tax returns do not generate any signals.
● There are 8,258 payroll deduction accounts that generated signals for a one-month period. If one supposes uniformity of the payroll deduction account signals over months, there would be almost 100,000 accounts signaled in a year. Note that it is likely that accounts would be signaled in more than one month and therefore there would be duplicates if one cumulated the signals.
● If all records signaled in a year are added, it gives the grand total of 244,269 signaled records. However, it is obvious that signals are duplicated between the administrative sources. For example, a change to the legal name of a business could be found on the tax return, as well as on each of its payroll deduction accounts.

TABLE 2
RESULTS OF SIMULATION STUDY

Adminis-trative Source	Number In The Profile Population	Number Signaled	Percentage Signaled
Individual tax returns	72,190	63,446	87.9
Corporation tax returns	102,688	81,727	79.6
Payroll deduction accounts	134,973	8,258	6.1

3.5 Questions Raised

The results of the simulation study, as well as an examination of the role of the signals, raise a certain

number of issues with respect to the profiling activities.

Six of these issues are presented below.

● **Performance of Signals in Detecting Change(s) to Statistical Entities**

The signals will attempt to flag legal and/or operating entities involved in real world events that have an impact on the statistical entities. An update will then be necessary on the central register to maintain the quality of the statistical products. Are the signals really reflecting real world events that affect the statistical entities or are there some that have no impact? If some are useless, work will be generated for no purpose.

A small-scale survey was conducted in 1986 to determine the usefulness of the signals with respect to the detection of changes to the statistical entities. However, for various reasons, the only signals that could be used were those of the simulation study. They refer to changes between tax returns of taxation years 1983 and 1984. Thus, the time lag between the reference period of the signals and the survey period (1986) gave recalling difficulties to the respondents. This led to the inclusion of events which took place after the period, as well as the omission of events which did occur in the reference period. The survey was therefore inconclusive and no other attempt has been made since then.

● **Repetitiveness of Signals**

Signals will be received over time and from different independent sources. The tax returns in particular suffer from noticeable time delays. As a given signal is received, the CFDB may have already been updated to reflect the real world event behind the signal. This update may have been the result of processing a signaled record from another source or of conducting cyclical profiling or of incorporating feedback received from surveys. Therefore, signals cannot be treated independently of the CFDB to decide to perform a reaction profile. However, how should a signal be checked against the CFDB to see if the CFDB was already updated? As an example, if a large increase in revenue is flagged on a corporation tax return, how should one check if the CFDB was already updated to reflect the real world event behind this increase, when one does not know the real world event behind it ?

● **Omission of Signals of Change**

Similarly, some records will not get signaled. Will the absence of signals definitively mean that no real world events occurred that need the statistical structure to be updated? Should other signals be developed to cover omissions? Again, the survey previously mentioned was inconclusive in answering these questions.

● **Availability of Resources to Handle Signaled Records**

As the simulation study showed, a large number of records will be signaled. These will require manual work. It is likely that there will not be sufficient resources to perform all this work. How should the total amount of resources to be devoted to reaction profiles be determined and how should this total amount be used to handle the signaled records? If constraints on resources demand that some signals be ignored, how will these be determined?

● **Response Burden**

The results of the simulation study suggest that businesses will be contacted more often than every

second year to check for frame changes other than through regular survey activity. This will increase response burden. Can a trade-off be established between increase in response burden and out-datedness of the register? What should this trade-off be?

● Role of Cyclical Profiling

The large amount of records signaled by the tax returns in the simulation study raises a question about the usefulness of cyclical profiling. The number of records subject to cyclical profiling and not to reaction profiling can be deduced to be very small. First, suppose the results of the simulation study in terms of numbers hold for a second year. Then, suppose the records signaled in the second year are not all the same in the first year but that there are new records signaled and that there are last year's records not signaled the second year. Then, it can be safely assumed that the number of records which will not get a signal over two years will be very small. There may be only a few records left which will not be signaled on either one or the other year. This will, in fact, represent the maximum target population for cyclical profiling. Will it be necessary to perform a profile for these entities, knowing that they are not signaled by the Payroll Deduction Accounts nor by the tax returns?

4.0 CONCLUSION

Section 2 has shown how administrative data were used to build a frame for initial profiling. Administrative data offered extensive coverage. However, it was also seen that conceptual differences between one's requirements and administrative data can lead to complications requiring simplifying assumptions and compromises.

The resulting frame supported the initial profiling of all business entities, the most complex ones. In these cases the approximation given by the frame could not be accepted. Rather, extensive research was conducted on each business entity using elements such as public annual reports and survey responses.

The frame also played an important role in initializing the CFDB. It was used along with the Business Register to identify the members of the Integrated Portion.

The method by which administrative data will be used to initiate a maintenance profile was described in Section 3. Signals of change will be derived from various administrative sources and will generate requests to verify profiles. Many issues were raised in this respect. These issues are being addressed by the various design teams responsible for implementing the CFDB update strategy. A solution being investigated to solve some issues is to prioritize signals depending for example on the length of time since the entity was last profiled. Another solution is to develop a self-learning process. Experience will dictate which signals are useful and should be kept. Therefore, substantial work is still required before the process stabilizes in production.

5.0 REFERENCES

Colledge, M. and Lussier, R. (1985), "A Strategy for the Provision of Frame Data and Use of Tax Data for Economic Surveys," Proceedings of the Section on Survey Research Methods, 1985, American Statistical Association, Washington, D.C.

Colledge, M. (1987), "The Business Survey Redesign Project - Implementation of a New Strategy at Statistics Canada," presented at the Bureau of the Census Third Annual Research Conference.

Statistics Canada (1987), "Version 4.2 of the CFDB Data Dictionary," Business Survey Redesign Project Working Paper, March 4, 1987.