

Re-thinking Statistics Canada's Business Register

Hélène Bérard, Stuart Pursey and Eric Rancourt

Statistics Canada

11th, Floor, R.H. Coats Building, Tunney's Pasture, Ottawa, Ontario Canada K1A 0T6

Helene.Berard@statcan.ca Stuart.Pursey@statcan.ca Eric.Rancourt@statcan.ca

Abstract

In the early 80's, major efforts were made at Statistics Canada to establish a central frame (commonly known today as the Business Register - BR) that could be used by most if not all economic surveys. Indeed, over the years, an increasing number of surveys have gradually been using the BR as a frame and it has become the backbone of the statistical programs of Statistics Canada's economic surveys. The methodology of the BR was designed to represent and organise as homogeneously as possible the complex legal and operating structures of businesses. However, since its inception, users of the BR have acquired a more in-depth knowledge of the structure of businesses. Further, with two decades of economic changes taking place, combined with more uniquely differing surveys that have been using the BR, and with a processing environment that has naturally aged, the time has come to re-think the purposes and structure of the BR. This paper presents Statistics Canada's Business Register and discusses the issues raised to adapt the BR concepts and structure to the current survey-taking reality.

Key Words: Operating Structure; Statistical Structure; Survey Universe File (SUF); Survey structure.

1. Background

Historical notes

The main pillar on which the survey sampling theory and practice rests is the existence of a finite list from which every unit can be identified and contacted. That's why people involved in the organisation and implementation of surveys have always devoted efforts in constructing and maintaining survey frames. To produce estimates about a given population, it must first be identified. This, in principle, is achieved each time a survey is carried out. However, in a centralised organisation like Statistics Canada where more than 300 surveys are conducted annually, it is unthinkable to construct and maintain a distinct survey frame for each.

In the context of economic surveys, businesses are often part of more than one survey. For instance, a manufacturing business could be part of surveys measuring either employment, investments, retail sales, wholesale sales, finance or exports. It is clear that the creation, organisation and maintenance of a common frame for most surveys - if not all - has numerous qualitative and financial advantages. That's why statistical agencies have invested in central frames for business surveys. At Statistics Canada, this frame is called the Business Register (BR). The BR was conceived and developed in the 1980's as part of a larger initiative called the Business Survey Redesign Project (Colledge, 1987). To create the BR, a standard survey structure was designed and implemented under the name of statistical structure. Combined with the organisation of collection activities through collecting entities, this has formed the backbone of the BR ever since.

At the end of the 1990's, Statistics Canada launched the Project to Improve Provincial and Economic Statistics (Statistics Canada, 1997) as part of an initiative to provide information to allow for the provincial re-distribution of the federally collected harmonised sales tax. This project had a significant impact on the amount of information to be collected from businesses and largely increased the number of BR users. Moreover, the use of new tax files and classification coding lead to further improvements (Castonguay and Monty, 2000).

Organisation of the paper

This paper is organised as follows: Section 2 describes the current Business Register and Section 3 presents the proposed changes that are part of the redesign. Then Section 4 raises some of the issues at stake and the conclusion follows.

2. The Current Business Register

General description

The BR is now supporting more than 90 surveys and is managed by a central unit (Business Register Division) which provides services to survey programs areas. The BR is made of a suite of files, programs and processes that interface with businesses through direct profiling, through survey responses and feedbacks and indirectly through administrative files (such as taxation files).

The BR is a list of live businesses that includes those engaged in the production of goods and services. Among these are both incorporated and unincorporated businesses except for some smaller entities¹. The BR covers all sectors of the Canadian economy whether businesses are involved in commercial, non-profit, religious, governments or institutional activities.

Construction and maintenance

The BR was originally constructed from specific survey frames and tax data. Today, it exists on its own and it is being updated on a continuous basis using three mechanisms. For large and complex businesses, updating is achieved through direct profiling which consists of contacting the business and establishing together its structure and contact points. This is a manual process conducted and maintained in the Business Register Division. For most businesses, the updating sources are administrative files produced by the Canada Revenue Agency (CRA). Among their legal obligations, enterprises must submit three sets of information to CRA: Goods and Services Tax collected (GST); payroll deductions (PD) retained from employees; and annual income tax forms.

The GST and PD files are obtained on a monthly basis and constitute prime information on the determination of presence of activity as well as the detection of new enterprises. They also provide information such as number of employees (PD) or taxable sales (GST) on the size of the enterprise. The annual income tax files provide a more detailed picture of each enterprise. In this case, two files are available: one for unincorporated businesses (T1) and one for incorporated businesses (T2).

Finally, when enterprises are contacted during the course of a survey, whether it be at the time of collection arrangements or at the actual collection time, any new information or change about the structure and classification of enterprises is fed back to the BR. This information is then used to update it.

Structures

As businesses can vary in structure complexity, there is a need to define a standard set of rules to adequately measure production units. Once the structure is established, various pieces of information are maintained. They include identification, location, contact information, structure, classification (North American Industrial Classification System – NAICS) and basic information such as the number of employees and/or gross business income. To officially function, businesses have a legal structure. To function, businesses have their own operating structure. To standardize our view of businesses, the BR maintains a statistical structure. All three structures are presented below.

Legal Structure

To exist, a business needs a legal status which implies a legal structure. Such a structure enables the business to communicate with government organisations such as the taxation agency (CRA). Each year, businesses must communicate with CRA using a unique Business Number² (BN) to report their income tax statement, to declare taxes collected as well as payroll deductions. These reports from CRA are transferred to STC under agreements and constitute the basis of the updating signals for the BR. The legal structure is maintained on the BR.

¹ Excluded are unincorporated businesses with no employees and with taxable sales lower than 30,000 Canadian dollars.

² Some self-employed companies are not required to have a BN.

Operating Structure

In their daily operations, businesses manage and organise themselves in a different way than the legal structure. Their structure is dependent on management methods related to the various business lines as well as the accounting practices. For instance, a single legal entity may be operating several plants and may own both wholesale and retail companies. These may all be under only one legal entity, but the operational structure could have two or more organisational and production units. The latter type constitutes the interest for surveys and the System of National Accounts.

Usually³, to be part of the BR, an operational unit should have employees, material, a manager and be recognized as an accounting unit. The type of accounting unit forms the basis of the operational structure on the BR. There are 5 types:

Investment Centre (IC): Responsible for accounting which provides profits and investments.

Profit Centre (PC): Responsible for specific revenues and costs.

Cost Recovery Centre (CRC): Recovers its costs by charging them to other centres for goods and/or services provided.

Cost Centre (CC): Unit for which costs are identified for management purposes.

Revenue Centre (CC): Unit which generates revenues⁴.

The operational structure is totally dependent on business cycles and management decisions. As a result, the concepts intended to be measured by the statistical agency may not be in line with the operational structure, hence the need for the creation of a structure with standard definitions of enterprise and production unit.

Statistical Structure

For sampling purposes, it is very desirable to have homogeneous units in order to have an efficient sampling design. But more importantly, to be able to correctly measure production as well as the flow of goods, services and capital it is mandatory to define standard units. A survey may be interested in gathering information on employment while another may be interested in financial statements. For these purposes, a common structure is needed. At Statistics Canada this structure is called the statistical structure. The statistical structure is a construction based on a series of rules to define and store on the BR a standard four-level structure hierarchy for all businesses. The highest level is the enterprise, while the lowest is the location. In between are the company and establishment levels. These are defined (Cuthill, 1990) as:

Enterprise: The enterprise is the highest level of the hierarchy and has a complete set of financial statements. This is the level where information about the international financial position is maintained. An enterprise may have one or more company.

Company: A level of somewhat homogeneous production with information related to balance sheets that allows derivation of the profit margin and return on investment. A company may have one or more establishment.

Establishment: Most homogeneous level in terms of production. It can provide information on total production output, cost of material, services and wages and salaries. An establishment may have one or more location.

Location: A unique physical production unit. The information available relates to employment.

A statistical structure is created for all businesses. Even in the context of simple one-level single-location production entities, all levels are created, but they all represent the same unit. For each of the lowest level units, an industrial activity code (NAICS) is assigned. Then, NAICS codes are assigned to higher levels based on a dominance rule.

³ There are exceptions. For example, an operational unit may not directly have employees, but could share them with other units.

⁴ It may have some marginal costs.

The enterprise and establishment levels are the main levels used for conducting business surveys. Based on these, enterprises are categorised as complex or simple in structure. A complex enterprise is one that is comprised of multiple establishments operating in different NAICS and/or different geographic areas. Conversely, a simple enterprise is one with a single establishment or with multiple establishments all involved in the same NAICS.

Interacting with surveys

While BRD is responsible to maintain the central frame, individual survey programs are responsible to define their respective needs. For each cycle (annual, quarterly, monthly) of a survey, a BR extraction is produced and the resulting file constitutes the sampling frame. Then the sample is selected and prepared for collection. Since the statistical structure is a construct by Statistics Canada, it may not always directly relate to an existing structure of the business capable of providing the requested information. As a result, collection arrangements need to be organised with reporting units. This may range from an aggregated report for multiple units to a series of reports at a finer level of details than required. In all cases, an allocation process is needed to distribute the collected information into the selected units in order to link responses with the concepts used by the survey.

All the information related to time-in-sample is kept and managed through sample control files. This allows survey methodologists to control the needed overlap (or non-overlap) between surveys to maintain the sample rotation strategies. Further, it provides the necessary information to reduce and manage the response burden of enterprises.

Finally, the survey information regarding contact persons, changes in the business structure as well as its size and presence or not of activity is then fed back to the BR for updating.

3. The Business Register Redesign

Need for a redesign

Since the 1980's, the economic situation has gone through many changes (e.g. globalisation, internet). Also, changes in tax policies and the advent of new administrative information have modified and improved the amount of information available to manage a central frame. More importantly, with close to two decades of experience in using the BR, numerous potential improvements have been identified. Also, with new integration needs (for instance for non-resident businesses for which we are interested in knowing the share of Canadian components), it is appropriate to review the practical and conceptual frameworks for the BR. Finally, the aging technological environment has been the catalyst in initiating the BR overhaul.

The objectives of the redesign are to i) simplify the BR; ii) increase the integration, harmonisation and coherence; iii) optimise the processes and allocated resources; and iv) update the technological environment and improve access. These objectives translate into changes in the concepts; changes in how survey structures will be defined; changes in interfaces between surveys and the BR and changes in the available information for designing surveys.

Concepts

The traditional approach to survey taking has been to first define a concept for then to attempt to gather information about this concept. In the case of business surveys, the traditional concept of survey unit has been materialized in the statistical structure. However, experience has shown that in many cases, no matter how the concept is defined, the level at which information can be made available does not change. Having realised this, great simplifications to the rules that define survey units have been identified. To implement them, definition of the survey structure will be performed directly from the operational structure (from what businesses can report). This may change concepts slightly, but refining them has been identified as an objective of the redesign.

Surveys using the BR are making use of standardised concepts as far as the structure, classification (industrial and provincial) and size are concerned. However, an internal review group has concluded that improvements could be made in how foreign businesses with Canadian control could be identified and down to which share percentage. Currently, three separate definitions exist and a common one could be maintained on the BR.

Change in survey structure

As seen in section 2, the current structure, the statistical structure, is derived from the operational structure. One of the major challenges in the redesign is to cease using the statistical structure and define survey units directly from operational units. Over the years, it has been noticed that in a number of cases, survey designers had been trying to get around the statistical structure and sampling not establishments for instance, but rather combinations of establishments, locations or groups of these. This suggests that even currently the concept of survey structure is defined for each survey by using rules which do not necessarily coincide with the a priori application of generic rules currently known as the statistical structure. By defining such rules directly from the operational structure, one intermediate layer can be dropped.

Directly using the operational structure has several advantages. Among these are the following (Gagné, 2004):

- The operational structure already exists on the BR and is fully maintained;
- The operational structure is based on standardized and homogeneous accounting rules;
- Often, the operational structure is simpler than the statistical structure; as a result the number of units to maintain would be smaller;
- The operational structure is more representative of businesses;
- Survey respondents better relate to the operational structure.

Making such a change is far from a drastic one. In fact, most businesses are simple enough in structure that whether the statistical or the operational structure be used as a basis for sampling, little or no change will result. Out of the 2.2 million businesses in Canada, less than one thousand have been identified as potential significant changes needing attention. Further, most of these are already part of the larger / more complex enterprises manually treated. Not having to maintain the four-level statistical structure for the other smaller / less complex businesses will be the main simplification.

Interfaces between surveys and the BR

The BR currently maintains a personalized set of interfaces with each BR-user survey. There is scope to standardise them. The main interface is the BR extraction that is produced as a snapshot of the population at a given time. This Survey Universe File (SUF) is then used for sampling purpose by the respective surveys. At present, it is produced from survey specific rules, but to simplify the process and increase coherence between surveys, it is planned to design a unique SUF framework that could be used by all surveys.

Defining the SUF is closely related to the concept of survey structure. Indeed, the survey structure can be defined by a series of rules from surveys, but it needs not be created a priori by the BR and maintained on it. In fact, all that is needed is to know how to use the operating structure through the rules defining the survey structure and directly create the SUF. This will have the coherence advantages and will avoid the need to create and maintain the statistical structure.

Design information

When the SUF is ready to be used for sampling, it contains a number of variables that can be used to elaborate the sampling design. The main two are the Gross Business Income (GBI) and the number of employees (EMP). They usually serve as primary measures of size when attempting to build efficient designs. However, these variables are often missing on the frame and must be estimated (modelled). Since variables used to model GBI and EMP are present on the BR, derivation of GBI and EMP is not really necessary. In keeping with the goal to simplify the BR, derived variables will no longer be produced and maintained on the BR. It will be the responsibilities of the respective survey methodologists to determine the best combination of variables to use in the design of their surveys.

4. Issues

Redesigning the BR raises several issues. Below are some of the more methodology-oriented challenges that Statistics Canada is facing.

Consistency between and within surveys

Currently, surveys are defined as establishment or enterprise surveys. When the survey frames (SUF's) are defined using survey structure rules, it will become more difficult to compare surveys in terms of their units if there is not a standardized set

of rules (or at least common guidelines). On the other hand, a survey interested in profit centres only will be more precise in this respect than one which is now using establishment as a proxy for profit centre. Though compromises will have to be made for some surveys, there will need to be flexibility to accommodate surveys targeting given levels of the operating structure.

Under the redesigned BR, surveys will either continue to survey the same level within businesses or they will slightly alter the survey structure to refine the definition of survey units. On their part, surveys not changing the structure will be compelled to try to simplify their rules to obtain a SUF. Surveys with significant changes will have to deal with potential breaks in data series and put in place a strategy to explain and/or adjust for it.

One or more model for the redesigned BR?

The current BR has a 4-level statistical structure created and maintained for every unit, whether it be a very complex or a one location enterprise. Given the dual nature of the structure (simple or complex), it could be interesting and perhaps simpler to have a BR for complex enterprises with a complete complex structure maintained and a simpler version for all the other simple units (with no structure). This option would give rise to two important issues: transition rules for migration between the two frames and capacity to expand complexity in case of economic, legal or political changes that would have a significant impact on business structures. Hence, the single model for all enterprises is favoured.

Tax data

During the last few years, we have seen significant increases in tax data usage in economic surveys. The BR redesign will face the challenge of making it a central data base capable of linking to all the various sources. In recent years, there have been demands to “know everything available about company X” to better design surveys and produce better statistics.

Updating mechanisms

Surveys are a major source of updates for the BR. When the use of tax data further increases, the number of feedback reports to the BR will decrease, leaving the updating process with fewer birth and death signals. To compensate for this, birth/death indicators should be refined, perhaps through modelling where appropriate. If the updating signals become too scarce, this may eventually justify the need for conducting a frame survey for the sole purpose of maintaining and updating the BR. There would be costs associated with such an activity, but also an increase in quality of frame information for surveys.

5. Conclusion

Redesigning a central survey frame such as the Business Register is a major endeavour. However, two decades of experience in using the current version of the BR combined with an aging technological environment have led us to conclude that the time is ripe to modernize it. One of the main issues is the switch from a statistical structure concept to that of a survey structure only maintained in the form of defining rules to create survey frames. As was seen, these changes (with the other ones planned) should improve the simplicity and consistency of the BR, and as a result, the design and quality of surveys.

References

Castonguay, E. and Monty, A. (2000). Recent Developments in the Statistics Canada Business Register. Proceedings of the Second International Conference on Establishment Surveys, American Statistical Association, 61-66.

Colledge, M. J. (1987). The Business Survey Redesign Project: Implementation of a New Strategy at Statistics Canada. Proceedings of the Third Annual Research Conference, Bureau of the Census, 550-576.

Cuthill, I. (1990). The Statistics Canada Business Register. Internal document, Informatics Branch, Statistics Canada. Revised 1997.

Gagné, P. (2004). Projet de Refonte du Registre des Entreprises, Internal document, Statistics Canada.

Statistics Canada (1997). Project to Improve Provincial and Economic Statistics, Catalogue No. 86N0003XPE, Statistics Canada.