

The Redesign of the Statistics Canada's Business Register

Yanick Beaucage, Paul Hunsberger and Stuart Pursey
Business Survey Methods Division, Statistics Canada

Abstract

Statistics Canada's Business Register (BR) is the central repository of information on businesses in Canada. The BR is used as the principal frame for the economic statistics program of Statistics Canada. Statistics Canada has undertaken a redesign of its BR, developed originally in the 1980s. An outline of the structure of the current Business Register is provided, followed by a presentation of the objectives of the redesign and details of the main elements of the redesign.

Keywords: Business Register, Redesign, Frame

1. Introduction

The Business Register (BR) of Statistics Canada serves as the frame for more than 90 business surveys at Statistics Canada. The foundation and concepts of the Business Register—the whole economic statistics program at Statistics Canada—are driven by the needs of the Canadian System of National Accounts.

The Business Register began to take shape in the early 1980s when efforts were made to establish a central frame that could be used by most if not all business surveys. Gradually over the years surveys have been linked to the BR—it is now the backbone of the business statistics programs of Statistics Canada.

Since the 1980s there have been enormous changes within and outside Statistics Canada. Fiscal policies, globalization, and the structure of businesses have changed. The conceptual basis behind the BR has become out of date. The revolution in informatics technology has made the informatics infrastructure of the BR difficult and expensive to maintain. New and useful administrative data sources have emerged. The functions of the BR have expanded over the years. There are more tasks and processes to maintain / update the BR. Survey methods have evolved and improved. Most importantly, with all these changes, user needs have evolved. The BR needs a rethinking—a redesign.

Section 2 of this paper provides an overview of the BR. Section 3 explains the need for a redesign with emphasis on the methodological aspects. There are three methodological issues: the creation of a sampling frame, the approach to updating the BR (measures of size and detecting births and deaths), and the development of a Quality Assurance strategy

2. The Business Register

The Business Register is a list of active Canadian businesses. The BR is the central frame for almost all of Statistics Canada's business surveys—and most especially the surveys that provide data for the System of National Accounts.

The BR has three major functions. The first and foremost is to serve as a frame for business surveys. The second is as a universe for the production of economic statistics using BR variables. The third is a database to help in respondent relations and the management of response burden on businesses.

The important BR variables for survey and sample design are the fields that contain data about geography, industry, employment, revenue, and data collection contact information.

There are about 2.29 million active businesses listed on the BR and 2.27 million of them are “simple” businesses. For these businesses, a single entity represents best their legal structure, operating structure, production unit, observation unit, sampling unit, and data collection unit. Although representing 99.1% of the businesses in Canada, these businesses represent only 40% of economic activity.

The other businesses are “complex”. There are only about 21,000 of them but they represent 60% of economic activity. The legal and operating structure of a complex business is represented by a set of entities (production units) arranged in a multi-level hierarchical pattern that illustrates the reporting relationships and data availability of the production units.

There are two useful views of a business. The Operating Structure is the business' view of itself—how it operates and structures itself. This structure is derived through a comprehensive interview with representatives of the business. This “profile” of the business is important to Statistics Canada because it identifies the types of production units within the complex business, the availability of types of data, and many measures required in survey design (size, geography, and industry).

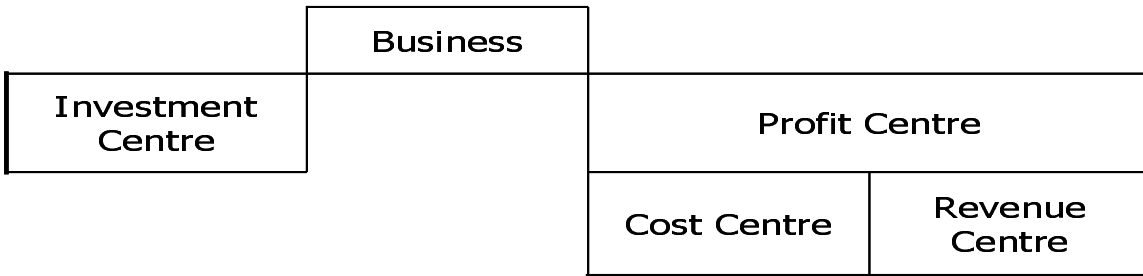
There are five types of production units in the Operating Structure. The Business Entity (BE)

represents all of the business and all data. The Investment Centre (IC) contains data on revenue, expenses and investment. The Profit Centre (PC) contains data on revenue and expenses—in effect “value-added” can be calculated. The Profit Centre resembles what we usually refer to as an establishment. The Cost Centre (CC) contains data on expenses only. The Sales / Revenue Centre (RC) contains data on revenue only. The diagram below illustrates the operating structure of a particular Complex business – it happens to have three levels separated into two branches, with several types of production units.

Figure 1a: a complex business



Figure 1b: Operating Structure of a complex business



The Operating Structure of a business forms the foundation to generate the Statistical Structure, using an algorithm called STATGEN. From this a sampling frame file for each business survey is generated. The Statistical Structure (Figure 2) is always a 4-level hierarchical representation of the business. Thus the Operating Structure of a business, whether simple (1-level) or complex (2 or more levels) is modified and expanded or squeezed into the 4-level Statistical Structure.

The top level of the Statistical Structure is the Enterprise and it is identical to the Business

Entity. The 2nd level is the Company (usually matched with the Investment Centre), the 3rd level is the Establishment (usually matched with the Profit Centre), and the 4th level is the Location (usually matched with Revenue Centres and Cost Centres).

The Statistical Structure is useful in sample design to generate sampling units. For example, in Establishment Survey we use the 3rd level of the Statistical Structure – the “establishment row”.

Figure 2: Statistical Structure of a complex business

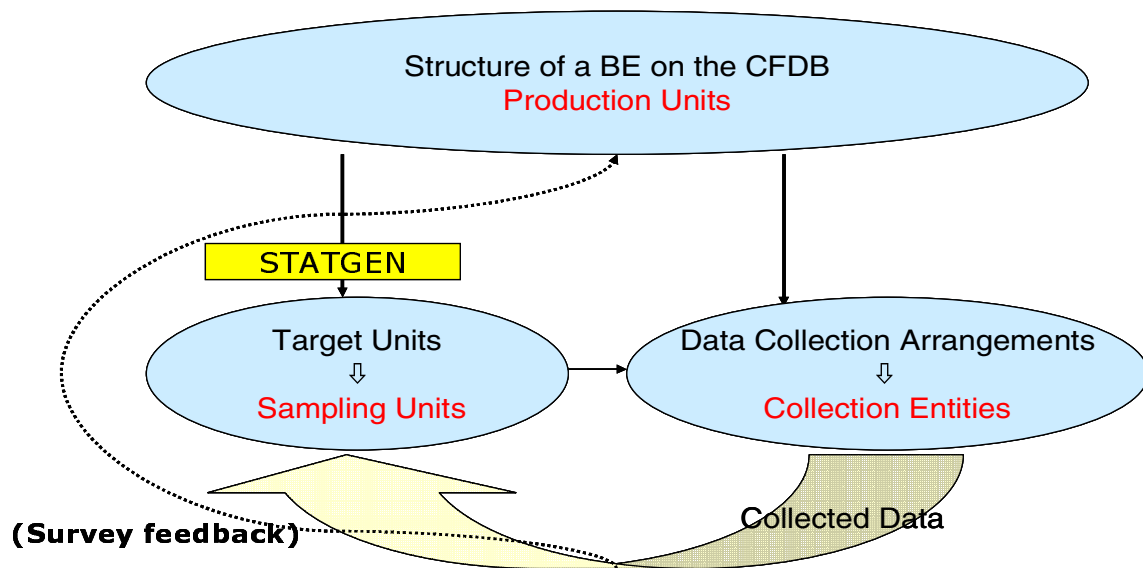
	Enterprise		
Company		Company	
Establishment		Establishment	
Location		Location	Location

The Statistical Structure has disadvantages – the transformation from the Operating Structure causes inconsistencies and artificialities in the business’ statistical structure.

The artificialities lead to difficult data collection arrangements—unnecessary response burden on businesses. Also the Statistical Structure database is expensive to maintain.

In the example, the branch on the right fits closely into the 4-level Statistical Structure. But the left branch requires modifications to get to the Statistical Structure.

Figure 3 below shows the structure of the BR—the flow of data and processes that take a business entity to a sampling unit to data collection.



3. Survey Design Issues behind the Redesign

The introduction to this paper emphasized the conceptual, economic, and technological reasons behind the need for a redesign. Within the redesign, an important goal is to simplify, optimize, and harmonize methods and processes of the BR. This section provides more detail on

the three major methodological issues of the redesign. They are: using the Operating Structure instead of the Statistical Structure in survey design, updating the BR (measures of size and the births and deaths of businesses), and developing a Quality Assurance strategy for the BR.

3.1 Using the Operating Structure

The Statistical Structure will be removed and instead the Operating Structure will be used to derive sampling frames. The Operating Structure of a business closely resembles the actual structure of the business. Business survey sampling frames will more accurately reflect the actual operating structure of the business. This structure is useful for designing data collection arrangements, in effect reducing response burden and getting accurate collected data.

The Statistical Structure, derived from the Operating Structure, is always a 4-level hierarchical arrangement, no matter how many or few levels the business actually has.

Its main advantage is that it makes sample design straightforward since the survey frame can be selected directly using one of the four levels. But the disadvantage is that the sampling units do not always match well with data collection units. As well, the Statistical Structure has inherent inconsistencies, sometimes inventing various units so as to maintain its hierarchical structure of enterprises, companies, establishments, and locations.

For example, in Figure 1b, if the Cost Centre and the Revenue Centre are from different provinces, then STATGEN creates a pseudo-establishment above each. This ensures—in the sample design—that economic activity is kept separate by province. But in data collection respondents will find it difficult to respond to an establishment based questionnaire. But in the example (in Figure 2) STATGEN created one establishment above the Cost Centre and the Revenue Centre – thus they were from the same province – and the establishment questionnaire representing the single establishment realistically reflects the respondent’s business.

Although the removal of the Statistical Structure will have significant advantages in data quality, response burden, and reducing costs overall, there is little impact on the flow of data (in Figure 3 the STATGEN process is removed).

As noted earlier, the 2.27 million simple businesses are best represented by a single entity, rather than the 4-level Statistical Structure. For the 2.27 million 1-level BEs clearly the BE is itself the sampling unit. The production units stand alone and unmodified

The remaining 21,000 businesses are complex (i.e. they are multi-level). Yet for all but about 450 of them, the development of an establishment survey is straight forward. This is because within each branch of a business’ Operating Structure, there is a level (somewhere) containing only Profit Centres. Taken together these Profit Centres cover completely, without duplication, the value-added economic activity of the branch. Then taken over all branches, the complete business is covered—that is, we are able to get value-added data for the whole business.

Although small in number, the 450 businesses represent a large 8% of the Canadian economy. These are the most difficult businesses to interpret—whether within the Operating Structure or the Statistical Structure. The reason for the difficulty is that there is at least one branch where there not a level, at the required detail of geography and industry, containing only Profit Centres. Instead, there is a Cost Centre or a Revenue Centre (that cannot provide value-added data).

As noted earlier, in the Statistical Structure this difficulty was handled in sample design by the creation of artificial establishments—which led to difficulties in data collection.

Thus it is ultimately the approach taken for these 450 businesses that will form the way that the Operating Structure is used as the foundation for developing sampling units in an Operating Structure Survey Universe File (osSUF). There are several options.

One option is to reduce the 3+ level businesses to 2 levels by calculating the dominant industry and province for each 2nd level production unit, based on the production units below it.

A second similar option, if the reduction to 2 levels loses too much detail, is to reduce these businesses so as to create a 3-level Operating Structure (thus all 3-level BEs require no calculations).

A third option is to create a 2-level Operating Structure, where in each branch, either the 2nd, 3rd, or 4th ... level is chosen to become the 2nd level of the Operating Structure. The choice of level is based on the amount of industry and geography detail that is required.

A fourth option is to travel down a branch until the lowest Profit Centre is reached. The production units along that row become the units in the osSUF.

Ultimately the choice will be based on simplicity of use, the popular requirement to do a value-added “Establishment Survey” (effectively a Profit Centre Survey), and the requirements of industry and province detail for the System of National Accounts.

Once the osSUF is created and used in sample design, each sampling unit derived from it is either a real production unit or a cluster of real production units from the Operating Structure. Each selected sampling unit points to its Business Entity, indicating that data will be collected from all or part of the BE. Since the sampling unit refers to a group of real production units, data collection from the BE (or part of it) will be smoother and easier.

The quality of data collection is critical in survey design. Thus it is quite possible that experience in data collection will provide feedback to help in fine tuning the structure of the osSUF.

3.2 Updating the BR

3.2.1 Measures of Size

One of the most important administrative data sources that became available relatively recently is a file based on the Canadian Goods and Services Tax (GST). The GST provides monthly data on business revenue and from it an annual measure of revenue size can be developed. This data source will be important in providing a timely measure of revenue size for the 2.27 million simple businesses. The traditional annual measure of revenue—T1 (for unincorporated businesses) and T2 (for incorporated businesses)—continues to be available. This annual source is not as timely as GST but is a valuable source to accompany the GST measure.

The Payroll Deduction Accounts file will be used as the source to provide direct data on the number of employees in a simple business.

However, for complex businesses, the profile interview will continue to be the primary data source for BR measures of size data.

3.2.2 Births

The BR loads, on average, 18,000 business births per month. Currently the BR loads births every three months. Occasionally there are constraints from administrative data sources. Also our internal process requires the assignment of industry coding at 6-digit NAICS (North American Industry Classification System), before business becomes a birth.

These constraints and process cause sudden peaks in the number of businesses on the Business Register and potentially sudden peaks in the estimates from business surveys.

Although it is difficult to manage the various sources of administrative data so that births are timely, we need to find ways to get around this difficulty and if not, find ways to lessen the impact of untimely births. One possibility is to accept industry coding at 3-digit or 4-digit NAICS, rather than the longer time required to code at 6-digit NAICS. This would most likely go a long way to allowing monthly updates of births.

Careful negotiation with players, with the goal of modifying processes within and outside Statistics Canada, may avoid the “avalanches” at less frequent intervals. Beyond that there is some consideration, if needed, of finding ways of smoothing the impact of births during estimation.

3.2.3 Deaths

The current rules to determine that a business is dead are conservative. It is easier to deal with businesses that are dead during estimation (using domain estimation) than it is to deal with businesses thought to be dead, but are alive (leads to under coverage).

But it is costly to maintain dead businesses on the BR, sample designs are not as efficient as they could be, and data collection costs are wasted. We need to find a balance between protecting against the bias due to under coverage yet keeping sample designs and data collection efficient.

There are several administrative data sources that are accurate for determining, with near certainty, the death of a business. These files will be

incorporated into the BR Redesign and a process created to “death” these businesses twice a year.

There is another type of business death. A business may be dying when it appears on a survey frame and be dead when data collection arrives. A business may be inactive (while administrative data files indicate that it is alive) and spring back to activity. In the BR Redesign we will explore the development of methods to manage businesses with inconclusive death signals.

Here are three examples of possible approaches. A survey process may decide to use a decision rule; based on the inconclusive signals to exclude a business from a survey frame, yet be kept alive on the BR. The rule would compare the risk of over-coverage to the risk of under-coverage. A survey may use a relatively inexpensive pre-contact before data collection to confirm the business’ dead / alive status. Other possible uses of the inconclusive signals are in the imputation strategy and in estimating the quality of the alive / dead status on the BR.

3.3 Quality Assurance Strategy

The third major methodological element of the redesign concerns Quality Assurance. In the redesign the goal is to develop a comprehensive Quality Assurance Plan that focuses on BR processes, measures of quality, and the important BR fields in survey design (measures of size, geography, and industry classification).

4. Conclusions and Timeframe

This section outlines the conclusions and the timeframe of the BR Redesign.

The Redesign of the Business Register is important to Statistics Canada, as almost all business surveys use the BR as the frame. Ultimately the data from our business surveys feeds the SNA.

Many redesign issues deal with the technical infrastructure of the BR. The major methodological issues deal with the creation of sampling units, updating the BR using administrative data sources (measures of size, business births, death) and the development of a Quality Assurance strategy for the BR.

The goal is to have the redesign completed and functional by 2008.

<u>Use of the Operating Structure</u>	
2005	Create new frames based on the Operating Structure
2006	Test the new frames with representative surveys
2007	Make modifications as required
<u>Methods for updating the BR</u>	
2005	Develop strategies for measures of size (revenue and employment numbers)
2006	Develop strategies for births and deaths
2007	Testing
<u>Quality Assurance Strategy</u>	
2006	Develop the strategy
2007	Development of methods and processes

Acknowledgements

The authors would like to acknowledge the work of two colleagues at Statistics Canada: Philippe Gagné and Julie Trépanier.

References

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, revised 1997). “The Statistics Canada Business Register” Internal document, Informatics Branch, Statistics Canada

Gagné, P. (2004). «Projet de refonte du registre des entreprises», internal document, Statistics Canada.

Bérard, H., P. Gagné, E. Rancourt, and S. Pursey (2005), «La Refonte Du Registre des Entreprises de Statistique Canada » Proceedings of the Colloque francophone sur les sondages, Québec, Canada

Rancourt, E., H. Bérard, and S. Pursey (2005), “Re-Thinking Statistics Canada’s Business Register”; Federal Committee on Statistical Methods Research Conference, Washington, DC