

RECENT DEVELOPMENTS IN THE REDESIGN OF STATISTICS CANADA'S BUSINESS REGISTER

Stuart Pursey¹

ABSTRACT

The Business Register of Statistics Canada was developed in the mid-1980s. Its role is to provide a frame for more than 100 recurring business surveys. Statistics Canada has undertaken to completely revise and update the infrastructure and approaches of the current Business Register. Methodological issues in survey design are discussed with emphasis on the proposed methods used to derive sampling units from the structure of a business. The new approaches in survey design are compared and contrasted with the current approaches.

KEY WORDS: Business Register, Business surveys.

RÉSUMÉ

Le Registre des entreprises de Statistique Canada a été développé au milieu des années 80. Il constitue une base de sondage utilisée de façon récurrente par plus de 100 enquêtes auprès des entreprises. Statistique Canada a entrepris une révision complète du Registre des entreprises actuel tout en mettant à jour son infrastructure et ses façons de faire. On discutera de questions méthodologiques reliées à la conception d'enquêtes en mettant l'accent sur de nouvelles méthodes proposées pour dériver les unités d'échantillonnage à partir de la structure des entreprises. Ces nouvelles approches seront comparées et on fera ressortir les contrastes entre celles-ci et les approches actuelles.

MOTS CLÉS : Enquêtes auprès des entreprises; Registre des entreprises.

1. INTRODUCTION

The Business Register (BR) of Statistics Canada is a structured list of businesses engaged in the production of goods and services in 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 purpose of a business survey is to measure the performance of the economy through the measurement of a) the production of goods and services in the economy and b) the financial position and performance of businesses. In business surveys the BR is used to create frames from which sample designs can be derived. The important BR variables for survey and sample design are the fields that contain data about geography, industry, employment, revenue and contact information.

The Business Register was developed during the 1980s with the objective of developing generic tools and frameworks for conducting business surveys; Cuthill (1990) and Colledge (1987) provide overviews. In April of 2005 the BR Redesign Project began. Gagné (2004) and Rancourt (2005) provide an early overview of this project. The fundamental purpose of the BR has not changed since the 1980s but there have been enormous changes in the environment surrounding the BR. Informatics technology has become hugely more capable, the BR is of increasing importance to STC (now serving as the frame for more than 100 surveys), new and useful administrative data sources have emerged (there is much greater use of administrative data at Statistics Canada), businesses operate in a more global economy, and there have been improvements and changes in statistical methods. In April of 2008 the “old” BR process will be turned off and the “new” BR process will be turned on.

This report compares the underlying approach of the two registers as they are used in the statistical methodology employed to design business surveys, and most particularly, their use in sample design.

¹ Stuart Pursey, Statistics Canada, 100 Tunney's Pasture Driveway Ottawa, K1A 0T6, Canada, Stuart.Pursey@statcan.ca

1.1 Overview of the Business Register

There are about 2,300,000 businesses on the BR. Almost all are simple businesses and their structure is represented by a single entity (or in a diagram by a single box). Almost 22,000 businesses are complex. Their Legal and Operating Structures, which are derived through interviews with the businesses, are represented by a set of entities (organizational and production units) arranged in a multi-level hierarchical pattern that illustrates the reporting relationships and data availability of the production units. Table 1 represents examples of two types of business in a diagram. It is the complex businesses where the challenging survey design issues arise. These businesses, though few in number, dominate the economy (Table 2).

Table 1 - Structure of a simple business and an example of a complex business

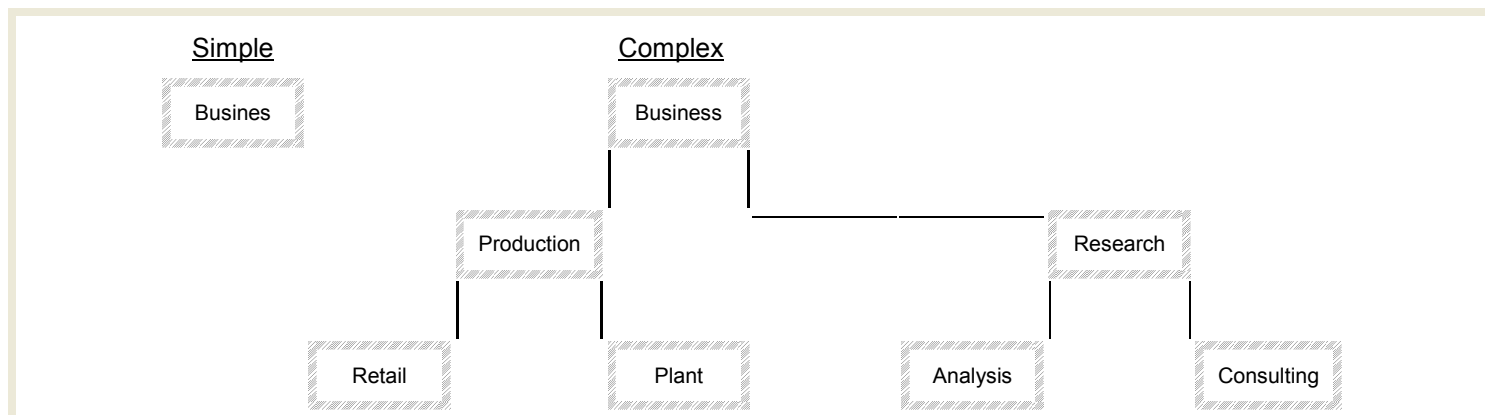


Table 2 - Counts and the distribution of revenue on the BR: Simple versus Complex by number of levels

<i>Type of business</i>	<i>Number of businesses</i>	<i>Count</i>	<i>Revenue</i>
1 level Simple	2,300,000	99.1%	40%
2 level Complex	20,000	0.9%	15%
3 level Complex	1,500	0.1%	16%
4 level Complex	250	Close to 0.0%	16%
>4 levels Complex	100	Close to 0.0%	13%

1.2 Issues in survey design as they relate to the use of the Business Register

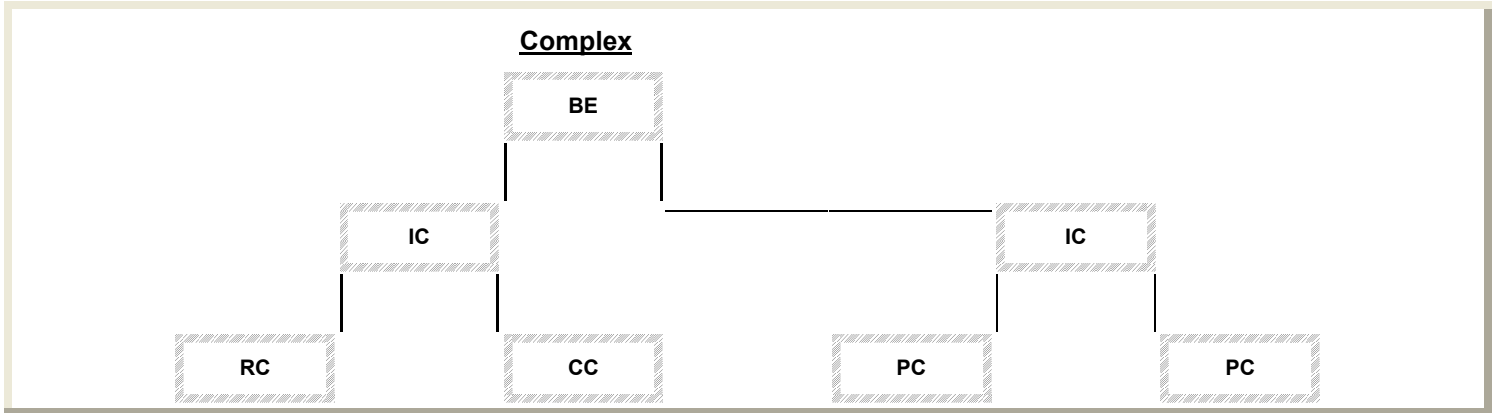
In using the Business Register, there are four major issues. First is the specification of the units (sampling elements) that will appear on the frame. For simple businesses the only option is the entity that represents the simple business itself. For complex businesses there are a myriad of choices (depending on the complexity of the business). In Table 1 the 3-level example of a complex business provides three possibilities: (the whole Business itself) or (the Production Unit and Research Unit) or (the Production Unit, Analysis Unit, and Consulting Unit) or (the Retail Unit, the Plant, and the Research Unit). Each of these options covers the whole business without duplication. But there are significant differences among these options. The determination of the “actual units” that cover completely without duplication the whole business is not a straight forward (Section 2).

The second issue concerns the variables in the BR that are required for stratification in sample design. Geography and industry variables are required because estimates are provided at these levels. Size variables are required to improve the efficiency of the sample design. The third issue concerns the identification of births and deaths of businesses quickly and accurately. For both of these issues administrative data plays an important role. The fourth issue concerns data quality – its measurement and ways of making metrics accessible to BR users. These issues are discussed in Bérard (2005), Gagné (2004), and Hunsberger (2005a and 2005b).

2. SAMPLING ELEMENTS

Table 3 shows the Operating Structure of the business in Table 1. Potentially a business might have up to five different types of units that are (usually) arranged as a hierarchy. At the top is the Business Entity (BE) representing all of the business, and able to provide all data. Underneath there is a selection of Investment Centres (IC) containing data on revenue, expenses and investment; Profit Centres (PC) containing data on revenue and expenses; and Cost Centres (CC) (data on expenses only) and Revenue Centres (RC) (data on revenue only).

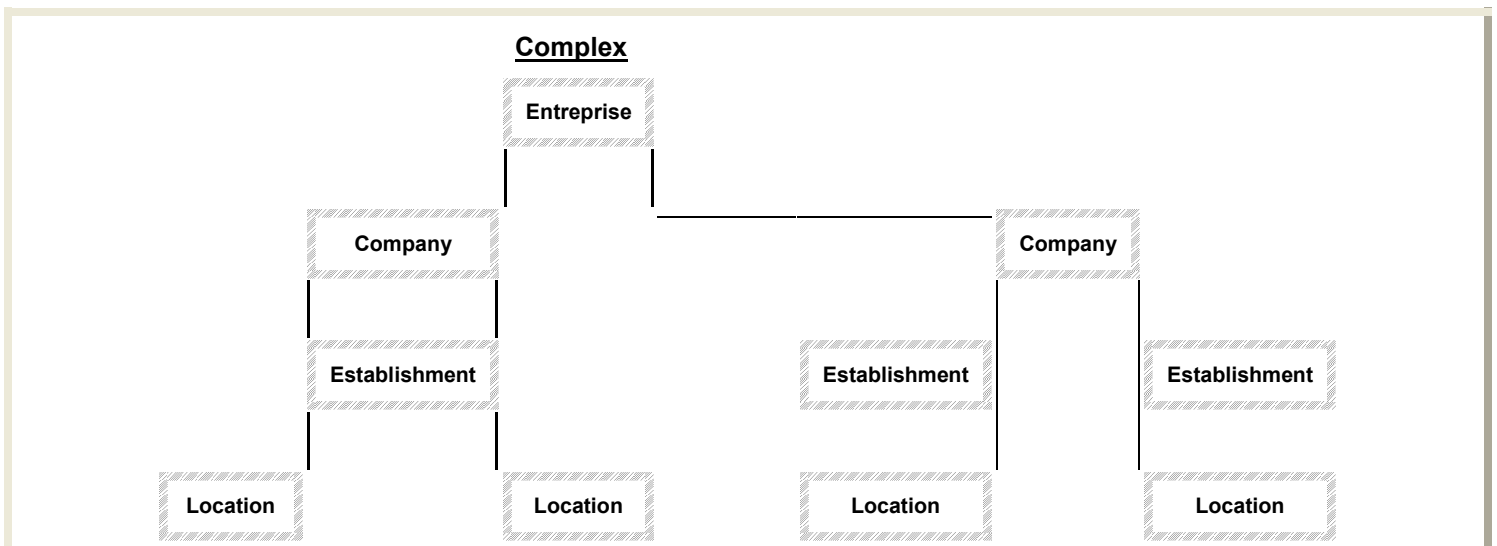
Table 3 - Example of a complex business



2.1 The current approach to deriving sampling elements

Currently the derivation of sampling elements is based on the Statistical Structure of a business. This structure is derived from the Operating Structure of the business using an algorithm called STATGEN. The Statistical Structure (Table 4) 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 a 4-level Statistical Structure. The purpose of this structure is to provide sampling elements for each of the four types of Statistics Canada business surveys: Enterprise, Company, Establishment, and Location surveys.

Table 4 - The Statistical Structure of the business in Table 3



The top of the Statistical Structure is the Enterprise: it is identical to the Business Entity. The 2nd row of the structure is the Company (usually matches to the Investment Centre), the 3rd row is the Establishment (usually matches to the Profit

Centre), and the 4th row is the Location (usually matches to Revenue Centres and Cost Centres). Note that in most structures, a unit in the Operating Structure must take on the multi roles in the Statistical Structure.

The best example of this is a simple business: that single unit in the Operating Structure must be replicated four times, once for each of the four Statistical Structure units. By inspecting and comparing Table 3 and Table 4, one can find the Operating units that play multiple roles in the Statistical Structure.

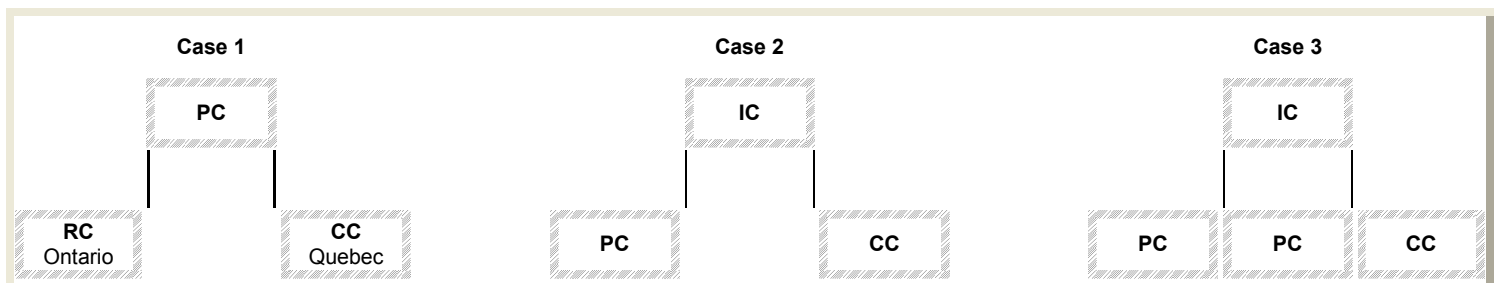
The disadvantage of the Statistical Structure is that the sampling elements do not always match well with the reality of the structure of the business. It has inherent inconsistencies, sometimes inventing various units so as to maintain its hierarchical structure of enterprises, companies, establishments, and locations. Note that the relatively straightforward 3 level complex business with 7 units (Table 3) has expanded into a 4 row business with 10 units (Table 4).

There are three particular difficulties with the Statistical Structure that occur in Establishment Surveys (see Table 5). They are best understood by realizing that STATGEN “wants” Profit Centres to become Establishments (and Investment Centres to become Companies and Revenue Centres and Cost Centres to become Locations.)

In Case 1, a PC has an RC and CC in different provinces. Rather than making the PC into a single Establishment (this would happen if the RC and CC were in the same province), STATGEN instead creates two “Provincial-split Pseudo-establishments”, one for each province. This is convenient in sample design since designs must be derived by province. But this causes difficulties in data collection: naturally the respondent will have difficulty in believing that its PC is two different entities.

In Case 2 the CC is a “sibling” of a PC. STATGEN realizes that the CC does not qualify as an Establishment and so it rolls the two together into a single “Roll-up Pseudo Establishment”. In Case 3 STAGEN does not know which PC to roll the CC into; so instead it creates the CC as an “Ancillary Pseudo Establishment”. Thus when the data collection units are created, sometimes they do not match the expectation of respondents; leading to confusion and unwanted response burden.

Table 5 – The special case examples



2.2 The approach to deriving sampling elements in the BR Redesign

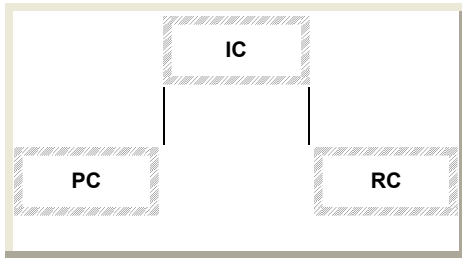
In the BR Redesign the Statistical Structure will disappear and instead sampling elements will be selected directly from the Operating Structure. The approach is based on understanding that if Location data are required, then the BE, IC, PC, RC and CC all are able to provide that data. The BE, ICs, and PCs are “over-qualified” of course! If Establishment data are required, then the BE, IC, and PC are able to provide that data (but not RCs and CCs). If Company data are required then the BE and IC are able to provide that data (but not PCs, RCs, nor CCs). If all Enterprise data are required, then only the BE itself is able to provide that data. Thus the strategy is to derive sampling elements as follows.

- In an Enterprise survey the sampling element is the BE itself.
- In a Company survey, Investments Centres are selected, with the condition that the set of ICs selected cover the whole of the BE without duplication. The strategy is: select the highest IC of each branch so that the whole BE is covered without duplication. If that condition cannot be met then select the whole BE.

- In an Establishment survey, Profit Centres are selected, with the condition that the set of PCs selected cover the whole of the BE without duplication. In Table 3, the IC in the left hand branch is selected and in the right hand branch the two PCs are selected. The strategy is: select the lowest PC in each branch so that the whole BE is covered without duplication. If that condition cannot be met, then units above the lowest PC must be selected instead. These are likely to be ICs but potentially, and at worst, the whole BE itself must be the sampling element.

How are Cases 1, 2, and 3 above handled? In Case 1 the PC is selected. There is no need to create “pseudo” sampling elements. Sample designs can still be derived by province because the information by province still exists (it resides in the RC and CC). In Case 2 and 3, the PCs are selected as sampling elements and also, the “sibling” CCs are selected as sampling elements. Thus there is no distortion of the Operating Structure. In sample design the information about the type of unit is available and thus the siblings CCs can be handled as desired in both sample design and data collection.

Table 6 – The case of an RC sibling



There is one special case in this strategy. When an RC is a sibling of a PC, instead of selecting both the PC and the RC as sampling elements, the unit above it (see Table 6) the IC becomes the sampling element.

This is done because operating structures of this type are best handled, for both sample design and data collection, by their “parent” unit. Again, there is no distortion of reality and data collection units can be set as desired.

- In a Location survey the sampling element is simply the set of all units at the bottom of the Operating Structure. In Table 3 these are the RC, CC and the two PCs.

3. CONCLUSIONS

In summary, the approach to deriving sampling elements directly from the Operating Structure has no disadvantages in the derivation of sample designs. Instead there are advantages that flow through the survey process, most especially improved respondent relations and reduction in response burden. As well, the removal of the Statistical Structure lessens costs since that structure need not be generated and maintained within the informatics system.

REFERENCES

- 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
- 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.
- Hunsberger, P., Y. Beaucage and S. Pursey (2005a), “*The Redesign of Statistics Canada’s Business Register*” Proceedings of Statistics Canada Symposium 2005, Methodological Challenges for Future Information Needs
- Hunsberger, P., Y. Beaucage and S. Pursey (2005b), “*The Redesign of Statistics Canada’s Business Register*” Proceedings of the Joint Statistical Meeting, Section on Government Statistics
- Rancourt, E., H. Bérard, and S. Pursey (2005), “*Re-Thinking Statistics Canada’s Business Register*”; Federal Committee on Statistical Methods Research Conference, Washington, DC