

# **Business Demographics from Statistics Canada: The Role of the Business Register**

## **Meeting of the group of Experts on Statistical Business Registers** organized jointly by UNECE, Eurostat and OECD

September 2015

Jamie Brunet

### **Abstract**

Since its inception in the 1970s, Statistics Canada's Statistical Business Register (SBR) has been used as a source for compiling statistics and analysis relating to business demographics. Various products have been developed to provide insight into the overall size of the Canadian business population, as well as show its industrial, geographic and firm-size composition. The SBR has also provided a basis for looking at the trends and the inner-dynamics of the population, including the provision of figures on business births, deaths and growth patterns. This aspect, however, requires additional data processing to transform the SBR enterprises records to improve their longitudinality. This paper looks at the effectiveness and continuing evolution of the SBR as a source for business demographics.

## **I. Introduction**

There are many important questions that are addressed through business demographic statistics. How many businesses are currently operating in a given industry, province or even city? How many new businesses are being created and in what sectors and geographic areas is firm creation most significant? How many firms are going out of business? How is firm formation and/or loss contributing to job creation or job loss? How many firms are growing or in decline? What are the characteristics of the fastest growing firms?

The Statistical Business Register (SBR) of a country is the base from which the statistics required to answer such questions can be compiled, but there are important considerations which must be taken into account in order for the SBR to do so effectively. In particular, the SBR must correctly identify the point in time at which the business becomes active or inactive and the data records to identify them should be longitudinal.

There are three basic components to this paper. First, the major business demographic data products are described and their strengths and challenges discussed. Second, the features of the SBR as it relates to business demographics are explored, including the related features of the tax data that are used as the foundation of the SBR. This includes a discussion on the potential for 'false' births and deaths shown by the tax data that must be identified and dealt with. Third, current efforts to enhance and further consolidate the processes required to produce business demographic statistics within the SBR production and maintenance framework are described, including options for future directions.

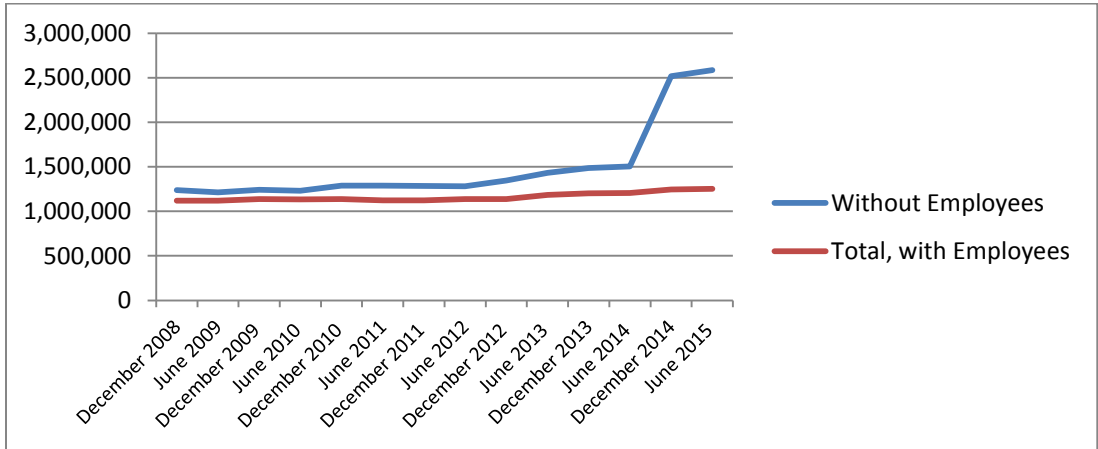
## II. Our Current Demographic Statistics

National and provincial counts of businesses from the SBR are published semi-annually. Known as *Canadian Business Counts*, this provides a timely snapshot of the number of all currently active businesses that either have recently paid employees or are incorporated. Unincorporated sole proprietorships and partnerships with no employees are also included, so long as their annual operating revenues exceed \$30,000. The product counts statistical enterprise locations and is published at the lowest level of the North American Industry Classification System (NAICS) maintained by the Register (6-digits). City level counts are also available by special request.

The *Canadian Business Counts* are useful as point-in-time estimates but cannot be used for the types of business demographic analyses that examine the dynamics and evolution of the business sector. There are two reasons.

First, there is no procedure in place to revise the published counts historically to take into account changes of a purely methodological nature. Improving the many processing rules and procedures that determine how the BR identifies and classifies active businesses is always on-going. In some cases modifications to these rules and procedures can cause significant increases or decreases in the *Canadian Business Counts*. Statistics Canada must therefore warn the users that the counts disseminated for one time period must be used very cautiously when comparing with those published previously. As an example, there was a major change in coverage for non-employer businesses that significantly increased the number of non-employer businesses for the December 2014 reference period.

Total Business Locations shown by *Canadian Business Counts*



Second, although generally speaking the SBR enterprise identifiers are often continuous through time, the SBR is not itself managed as a longitudinal database. As a result, it cannot be said that the creation of a new enterprise record as represented by a newly-existing identifier signals a true business start-up, nor that the disappearance of a given enterprise identifier signals a true business closure. This hinders the derivation of

accurate figures for the purpose of analyzing firm entry, exit and life-cycles. It can also cause distortions in the period-to-period net changes shown by the Counts.

To overcome the limitations of what the SBR can provide to support demographic analysis, Statistics Canada has developed and continues to maintain other datasets and programs that are based on the SBR.

There is the *Longitudinal Analysis Employment Program (LEAP)*, which was first created in the 80s and continues to be updated on an annual basis. It is a longitudinal data set of enterprises that provides a basis for tracking how employment change relates to business demographic events. One of its key functions is to measure to what extent business start-ups are driving employment gains and, conversely how business closures are driving employment declines nationally and also within industries and provinces.

Over the years, Statistics Canada's has used the LEAP as the basis for various expanded longitudinal datasets and research projects that include key economic variables such as business revenues, other financial variables from administrative records and in some cases survey data.

The LEAP 'longitudinalizes' enterprise records from the SBR by creating a separate longitudinal enterprise identifier so that the appearance or disappearance of an enterprise on the dataset is more reflective of a true firm entry or exit as opposed to an administrative event or a change in the enterprise's status. Year-over-year pairs of enterprises from the SBR are compared against one another to find distinctly identified records that should be joined into the same longitudinal record. LEAP data are therefore not be affected by the false entries and exits that would be present from the SBR records themselves.

This process is partly conducted using automated and manual routines comparing enterprise names and addresses and other information. Additionally, a key feature of the LEAP is that the employees that are associated with an enterprise are also used to identify false entries and exits. The end-of-year tax slips (known as T4 slips) sent to all employees summarizing the earnings and deductions made are used to compile the list of workers belonging to a business in a given year and searching for a similar cluster of workers among other employer records in the prior or subsequent year. As a result of this 'labour tracking', some of the newly-created or discontinued enterprise identifiers on the SBR are then determined to belong to businesses that were in fact continuers and not new or deceased ones as the SBR would initially indicate

In more recent years Statistics Canada has also produced annual *Entrepreneurship Indicators*, which are also based on the SBR and, like the LEAP, provide a basis for analyzing the demographics of employer businesses, but with a slightly different definitional and procedural framework. The focus of the *Indicators* has been to provide enterprise births and deaths that are comparable across OECD countries, according to the definitions in the *OECD-Eurostat Manual on Business Demographic Statistics (2007)*.

These data examine the employment size of births and their subsequent survival rates, and identifies high growth firms in terms of employment and revenue. Following the Eurostat-OECD guidelines, enterprise births (deaths) amount to the creation (dissolution) of "a combination of production factors with the restriction that no other enterprises are involved in the event". That is "births and deaths do not include changes in the population due to mergers, acquisitions, break-ups, split-off or restructuring of a set of enterprises". Births exclude newly created enterprises that simply assume the activity of an existing enterprise (take-over); similarly, deaths exclude enterprises subject to a take-over. Both births and deaths exclude reactivated enterprises that restart activities within two years. In order to ensure that the statistics

produced comply with the Eurostat-OECD guidelines, a number of different filters based on variables available in the BR are employed to identify false births and deaths. These filters are supplemented by a subject-matter review.

### **Business demographic statistics at a crossroads**

Statistics Canada has developed and continues to maintain these business-demographic products and data sets, as these meet many of the needs of users. There are, however some key challenges that need addressing.

The *Canadian Business Counts* are a very current source of statistics with extensive industrial and geographic detail. But it does not provide gross-change figures on firm and entry and exit and the net changes it shows are often not indicative of “true” changes.

The ‘longitudinalized’ datasets and products used for the purposes of the LEAP and *Entrepreneurship Indicators* provide a solid basis for gross-change figures and firm-dynamic analysis, but they cannot be used to assess *current* economic trends, due to the two-year time lag required for production of the data. This is in part due to time required for the data development itself, and also due to the amount of time it takes for some of the data inputs to be made available (for example, the T4 slips cannot be provided by the tax office until a full year after the reference year.)

Business demographic statistics and analysis are clearly important. We are commonly asked for information regarding level of firm formation and disappearance. Often, policy researchers or business planners want statistics to tell them where the ‘hot spots’ are, i.e. in what sectors and/or areas of the country are new business ventures being created ? Often, the desire is to know for very specific local areas.

On a broader scale, there is keen interest to understand how the entrepreneurialism is driving trends in the macro-economy. In recent public speeches, the Governor of the Bank of Canada has pointed to the importance of knowing to what extent the creation of new businesses is creating jobs. The data sources discussed above cannot quantify the relationship on a current basis. Economic analysts and data users want and need a far more up-to-date set of statistics to help understand these phenomena.

As a means to provide data on the creation of new employer businesses and the job creation associated with it, Statistics Canada is working on experimental estimation methods that would use past trends in false births and deaths as observed through the LEAP and use them to produce more timely estimates on a quarterly basis. These experimental estimates are currently available and being used by the Bank of Canada.

Statistics Canada has formed an internal working group consisting of individuals from the divisions that have a role to play to determine how best to meet the challenges. Its focus is to first assess and explain the technical and methodological differences between the different data products, identify areas for improvement and then explore means of addressing them. The SBR is and will remain at the heart of business demographic statistics. Our division is therefore a key part of this working group and is an integral part of this process. We are especially focused on assessing the longitudinality of the BR and finding ways to enhance it.

## **II. The SBR: A Solid Foundation for Business Demographics**

The central role of the SBR in these data products implies that they are based on the SBR's concepts, procedures and data sources for identifying and counting active businesses and tracking them through time. The SBR provides a complete listing of the business population that is built and updated from a rich supply of tax data. There is also extensive updating through a large team of statistical enterprise profilers and also through feedback received from respondents in the course of conducting the agency's business surveys. It is therefore a very capable resource for business demographic statistics.

At its core the SBR, is a compilation of statistical enterprises identified through data received from the national tax office, the Canada Revenue Agency(CRA). There are currently some 6 million active legal enterprises on the base.

About 60% of these are identified through the CRA's own business registration program, which assigns a unique identifier -- known as the Business Number (BN) -- to each business to be used for the purpose of collecting many types of taxes. There are three types of BN-based taxes that form foundational elements of the SBR:

- 1) corporate taxes,
- 2) employee payroll deduction remittances and T4 Statement of Remuneration Paid,
- 3) goods and services sales taxes

In Canada, all legal corporations must file an annual corporate tax return; all employers must deduct income taxes from employee's paycheques, and most businesses earning more than \$30,000 in annual revenues must collect and remit goods and services sales taxes (a kind of value-added sales tax). Having to pay or remit any one of these requires that the business apply for a BN from the CRA. Collection of these taxes is then managed by the CRA using the BN as a centralized account number for the business.

The BN provides a very reliable basis for bringing different types of tax data together for the purposes of building an unduplicated and up-to-date listing of enterprises for statistical purposes. The BNs tend to be consistent through time (ie businesses tend to keep their BN identifiers from one year to the next) and conceptually the legal entities they represent tend to correspond well to the definition of a statistical enterprise, (though in some cases, several BNs can be grouped into a consolidation which would in turn correspond to a single autonomous statistical enterprise.)

The remaining 40% of the enterprises on the SBR are identified using the personal income tax returns of unincorporated sole proprietors and partners who do not indicate a link to an existing BN through which employee payroll deductions or value-added taxes would be remitted. (Corporate taxes would not be paid by these individuals by definition.) The personal identifiers used for these unincorporated businesses are kept from year-to-year. Like the BNs, they provide a solid foundation for longitudinal records on the SBR.

The continuity of business identifiers through time is one important aspect of business demographics. Being able to determine when businesses become active or inactive is another. The tax data generally do this effectively. Statistics Canada receives updated information for the tax reports and remittances it uses on a monthly basis from the CRA. These data are used to continuously update the SBR's Business Status variable, and also the estimated total employment and annual revenues, which are also maintained for each and every enterprise on the SBR. The business status and employment and revenue variables form the basis

for identifying which enterprises are contributing to economic production and should therefore be counted as an active business for the purposes of business demographics.

To ascertain whether the business is currently earning revenues, we check for income tax revenues from the corporate tax returns or, in the case of unincorporated sole proprietorships, from the personal income tax forms. Corporate tax returns must be filed for each fiscal year, within six months of the period end-date. Late submissions are subject to penalties levied as a percentage of payable taxes. Personal income taxes for sole proprietorships and partnerships must similarly be paid soon after the end of the reporting year or face penalties.

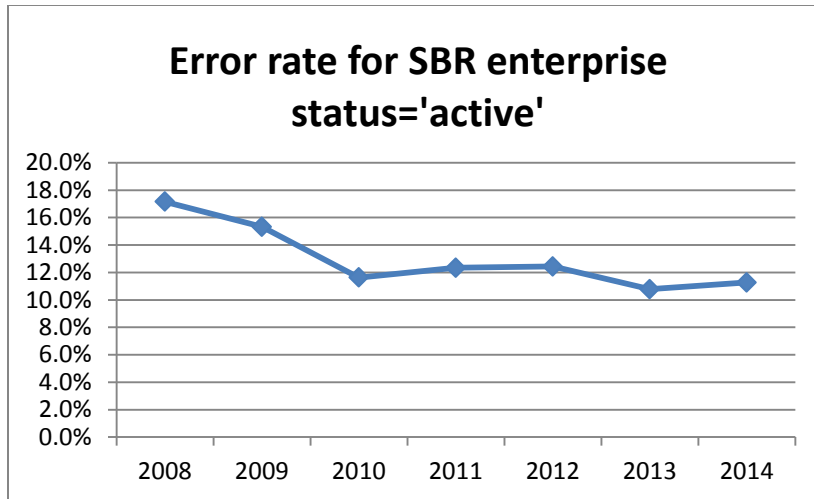
The goods and services tax can also be used to obtain revenue amounts. These taxes have the added benefit of a requirement to pay quarterly or monthly. Only the smaller businesses are permitted to pay annually.

Payroll deduction remittances tend to be very effective as a source for determining current employment levels. This is due to the fact that virtually all employers must provide employee payroll deduction remittances at least on a monthly basis. Moreover, when providing these remittances, the businesses are also asked to indicate the total number of employees for whom payroll deductions have been made. From these payroll deduction data, Statistics Canada obtains an accurate and up-to-date estimate of employment for every business in Canada. In fact, these data form the basis of the agency's official total payroll employment figures disseminated through the labour statistics program.

The income tax revenue, taxable sales revenues and total employment are used in conjunction by the SBR to determine the life status of the business. We assign a value of 'active' to any business for which recent non-zero values in these variables have been observed. Given the reporting requirements for tax purposes, the methods used are considered very effective for identifying business start-ups, though there are circumstances that can indicate 'false start-ups' (see below).

Realistically, however, identifying that a business has 'died' and the precise moment of its death is more problematic. There is a process by which a business should close its BN once it has ceased operations, but there is generally no formal requirement to do so. Moreover, our research and past experience have indicated that activity can re-emerge even after a prolonged absence of corporate taxes, paid employment or value-added sales tax revenues. This may be due to late filing or periods of dormancy. As a result, as a practice Statistics Canada avoids flipping the status of a business to 'inactive' prematurely. The algorithm for making this switch is very complex, but in essence, there must be no observable payroll within the last 13 months AND no observable value-added tax remittance with last 16-25 months, AND no observable corporate or personal business taxes paid within the last 3 years.

These rules have been refined in recent years with positive results. Quality assurance metrics compiled through sample contacts to enterprises on the SBR have shown a significant improvement in the error rate for 'active' status category, declining from 17.2% in 2008 when the rules were first implemented to 11.3% as they have been refined. As usual, the error rate for larger businesses is much lower, as evidenced by the fact that the revenue-weighted error rate is well below 1%.



Accurate measurement of business deaths is a normal challenge. As per the Eurostat-OECD guidelines on Entrepreneurship, there are complications in identifying the effective date of a business death. The situation in Canada is no different.

As a final note on assigning business activity status, it is worth mentioning that we also benefit from survey feedback processes whereby respondents to business surveys are asked to update the key frame variables, including business names, contact information and the current operational status. This information is then used to update the Register. This provides the opportunity to correct the business status in many cases, as this procedure is a built-in part of Statistics Canada's many survey programs.

Our enterprise profilers on the SBR team and also in the Enterprise Portfolio Manager program also play a key role in keeping the business status variable up to date, especially for larger enterprises.

### **The potential for false births and false deaths**

Business births and deaths are typically compiled from the SBR by comparing the active enterprises listed in a given time period  $t$  with the list of enterprises that were active in a previous time period  $t-1$ . Those present on the list for  $t$  but not  $t-1$  are considered *potential* births. The converse implies a *potential* death.

Given this approach, the SBR enterprise identifiers would ideally continue longitudinally in a manner that is in line with the definitional framework required for business demographic statistics. In some situations, however, this is not the case.

As stated above, the creation and dissolution of statistical enterprises on the SBR is, for the most part, based on the tax data. The automated processes of the SBR typically create new enterprise identifiers whenever there is a newly active CRA business number, or in the case of sole proprietors that are non-employers with no taxable goods and services sales revenues (and therefore no BN), a newly-identified individual reporting business-related income.

There are circumstances under which a pre-existing business or some part thereof is designated a new BN by the CRA. Some key examples are as follows.

- The decision by a sole proprietor to incorporate his or her business constitutes a **change in legal type** and will therefore necessarily require the creation of a new BN account. Any BN that would have been used by the business as a sole proprietor (for payroll deduction remittances or value-added taxes) would need to be closed. In many cases, the new corporation will have a name with resemblance to the operating name of sole proprietorship, but in some cases it will not.
- Corporations can amalgamate with one another, often requiring the termination of the existing BN accounts and creation of a new one to represent the amalgamated entity. The term “amalgamation” is the formal term used by the CRA for situations where businesses join together under the ownership and control of the corporate directors or owners. Fortunately, the CRA is required to keep track of the links between the discontinued BNs that are amalgamated and the new BN that is created. The CRA provides a link file to Statistics Canada that is used to populate a “successor” field on the SBR. For an amalgamated enterprise, this field will list the identifiers from which it was amalgamated that have been closed.
- A change in ownership or the directors will also sometimes require a new BN, under specific criteria.

### **III. Current and future development work**

Much of the work to be done in the coming months is to look more closely at the nature of birth and death records on the SBR and assess the effectiveness of the various methods in place to identify and correct them.

Currently, these corrections do not impact significantly on the overall birth and death rates. For example, the published LEAP statistics indicate a total of 137,500 entrant businesses (private-sector only) for 2013. Relative to the overall population of private employer businesses of approximately 1,090,500, this would yield a ‘birth rate’ of 12.6%. For a given reference year, the LEAP methodology typically reveals about 6,000-7,000 enterprise identifiers from the SBR that should be ‘paired’ with a different enterprise identifier for the preceding or subsequent year. The relative impact of false births and deaths on overall entry and exit counts is therefore not large. The impact on the estimates showing job creation and job loss as a result of demographic events is much larger. Since this is a key aspect of the statistics to be produced, correcting for false events is essential.

One question to be answered is whether the procedures in place are finding them all. We are also focused on automating as much of this work as possible. Much of it currently entails time-consuming manual review. As a result only the largest enterprises can be looked at.

The processes make significant use of the amalgamation links made available by the CRA, but there are only about 200 of these that are received per month. Automated searches on the business operating names are also used. These searches use parsed versions of the names and compute a score to determine the likelihood that a matched pair is indicative of a false birth or death. Improved name parsing and score functions are currently being tested.

Another important focus for the SBR team is to identify means of incorporating information required for longitudinal analysis and statistics on to the Register itself. Currently, the LEAP and Entrepreneurship Indicators datasets are built from the SBR but are currently maintained as separate datasets. The identified



false births and deaths are not fed back to the SBR and the SBR itself continues to be maintained in a manner that longitudinal data cannot be derived *directly* from it.

A key objective of the working group that has been established is therefore to identify options and recommend a means of consolidating the methodologies and practices related to business demographics and incorporate them into SBR system and procedures. This is seen as key to the continued enhancement of the coherence and usefulness of the resulting statistics.

To do this, as per Rollin (2012), we hope to build a taxonomy of firm changes name, address, industry, employment, ownership so that firms can be linked over time in a variety of ways depending on the analytical need.

There are some different options that can be considered. We can build upon variables already in the SBR system that identify ‘successor-predecessor’ links between enterprise identifiers across time periods . These variables currently show the relationships identified by the data on amalgamations received from the tax office. We are exploring the option of building upon these tables so that they would also show the relationships discovered by the false births and deaths identified through other processes.

There is also the possibility of building a longitudinal enterprise identifier on the SBR which would be distinct from the main enterprise identifier. This would facilitate the compilation of longitudinal business statistics for various purposes. It would ensure consistency in concepts and procedures, thereby optimizing efficiency and effectiveness. A challenge, however, would be how to handle the historical SBR snapshot data. Currently our SBR processes create a new snapshot each month. The snapshots for previous snapshots are not recreated, because it is not necessary given the purpose of providing survey sampling frames, which is the primary purpose for which the SBR has been built. Longitudinality implies that links between enterprises in the historical snapshots are discovered subsequent to the point at which they were created. For example, it may be learned that distinct enterprises identified previously have been linked by way of a merger. This may create the need for separate data files of longitudinal links within the SBR system that can be re run historically, which would be a significant change to our operations.

The working group has been mandated to look at the many technical aspects and recommend a strategy by the end of 2015. The next re-design of the SBR system is scheduled to begin in 2018 (for completion in 2023). Depending on the nature of the recommendations, the changes to be made may be implemented prior to the re-design, or may be planned as part of it.

## References

Canada Revenue Agency, general information for businesses, <http://www.cra-arc.gc.ca/bsnsss/menu-eng.html>, 2015

Eurostat-OECD (2007). *Eurostat-OECD Manual on Business Demography Statistics*, <http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf>

Eurostat-OECD (annual), *Entrepreneurship at a Glance*.

Rollin, Anne-Marie, *Developing a Longitudinal Structure for the National Accounts Longitudinal Microdata File (NALMF)*, Proceedings of Statistics Canada Symposium 2013

Statistics Canada, *Business and employment dynamics data at the national level, 2011*, *The Daily*, June 30, 2015 (Includes links to data tables compiled from the LEAP and documentation on data sources and methods)

Statistics Canada, *Canadian Business Counts, June 2016*, *The Daily*, August 6, 2015 (Includes data release and links to tables and documentation on data sources and methods)

Statistics Canada, *Entrepreneurship indicators of Canadian enterprises, 2011 and 2012*, *The Daily*, March 5<sup>th</sup> 2015 (Includes data release and links to tables and documentation on data sources and methods)