



Group of experts on Business Registers
27 - 29 September 2017

How can data collection improve? – An outlook from the Swedish Business Register



Introduction

The ultimate goal for all development work in the Business Register (BR) is to increase the quality. In a broad sense this is what is comprised in the principles of the Code of Practice together with the general quality management principles in the European Statistical System. Another way to put it is to try to help decrease the total error in the statistics that make use of the BR. For the Swedish BR that have been developed over four decades and with the benefit of institutionalized cooperation between governmental authorities – how can this be done? Given the specific circumstances surrounding the organisation – for instance legal, financial and technical – some options come out as more likely to give effect than others. For the Swedish BR the efforts in recent years have been directed to technical development for increasing productivity in the production of the BR. A reengineering project is currently running. Tools for editing and data collection are outcomes so far. Coming steps would be to move to web based techniques that affect data collection to a larger extent than today.

Background

In 1963, the Swedish Parliament commissioned Statistics Sweden to maintain a central business register. The purpose was to provide business statistics with the conditions to meet the requirements of degree of coverage, coordination, timeliness etc., and consolidate the various business registers maintained at other departments at Statistics Sweden. The central business register was initially updated twice per year with information from the tax administrative registers.

In the 1960s, technology was developed to create comparable statistics through the use of coordinated frame populations and sampling. In 1972, the system for coordinated sampling (SAMU) was applied for the first time. An annual version of SAMU was created in November, and extra efforts were made to ensure that the quality of the Business Register would be the best possible at that time. From 1975, information was collected from the 25 000 largest enterprises via an annual survey. In the beginning of the 1980s, the Business Register was supplemented with information from public sector local units.

An extensive collaboration with municipalities to improve the quality of the local unit segment was carried out in conjunction with the 1980 Population and Housing Census. After these efforts, the Business Register was considered nearly complete by the early 1980s, even including companies' local units, and it was renamed the Central Business and Local unit Register. Through the work that was carried out in the Base register study, which resulted in the Ordinance on the Public Business Register in 1984, conditions improved significantly

for the maintenance of the Business Register by means of the prescribed right to utilise a number of other administrative records. The ordinance defines the part of the Business Register that was the public business register. It contains fewer variables, and a simpler enterprise concept compared with today's Business Register, where an enterprise corresponds to the legal unit and its local units. Approval was not required for distributing information in the public business register in Sweden.

Once Sweden joined the European Union (EU), an adjustment was made to the EU regulations on economic statistics and business registers for statistical purposes. Thus the contents of the business register have been expanded with several new types of units: Enterprise, Kind of Activity and Local Kind of Activity. Information on relations between legal units makes it possible to form Enterprise Groups for the EuroGroups Register (EGR). The current system became operational at the end of 1999. The Business Register has been renamed to reflect this expanded content and is now called the Business Register.

Purpose and legal framework

The BR has since 1984 two main purposes: to serve as a register for the production of statistical sample frames and to serve as a common source for basic information on legal and local units for public use. The 1984 regulation regarding the public business register is worth mentioning a little extra. The most important features of this regulation are that it stipulates in detail:

- What kind of variables and units that shall be included
- Which government agencies that shall supply the data
- When shall be delivered to Statistics Sweden

The result is a register based on weekly updates on legal units and a number of related variables. The main data suppliers in this regulation are the Tax Authority and the Swedish Companies Registrations Office.

In effect, the requirement for a virtually constantly updated register has forced the BR to look for data sources that can supply data fast and with short or preferably no time lags. This is a challenge. The array of data sources available and the quality of the data reflect to some extent the legal context for a statistical bureau. Unlike in some countries in the EU-community, it is not compulsory for legal units to continuously register and deregister local units, and likewise, there is no coordination at national level of valid visiting addresses for business locations.

Major sources

The Tax Agency and the Swedish Companies Registrations Office deliver basic information for most legal units on a weekly basis. In addition to this, a number of other administrative records based on the tax regulation, such as gross wages, income statements, tax deductions and turnover is sent to Statistics Sweden on different frequencies. Information on other legal forms not covered by the government agencies mentioned above is also provided to the BR.

In order to update information for local units a yearly survey is directed to all enterprises with more than one local unit in the BR at the time of sampling in August. In total around 8 000 legal units with about 88 000 local units are surveyed.

Complementary data sources

Over the years an intricate system of data sources has been established to feed the BR. Each source serve different sets of variables in the register. Some sources even contribute with updates for one variable only, like if a company is registered on the stock exchange or not.

An overwhelming majority of sources supply data on legal units. Some sources are used to complement each other. For instance, postal addresses are supplied both from the Tax Agency and a commercial source. The reason is that the commercial source also provide information on addresses for local units that are present in the BR. These cannot be retrieved from the Tax Agency which only need information on postal addresses for legal units in order to carry out its duties.

Commercial data sources have become increasingly important over the years, not the least for supplying information for the creation of enterprise groups.

Implications for the BR-system

The BR-system of today has to be able to cope with how to edit all data on a weekly basis. Once Statistics Sweden is notified, a large number of checks are made concerning data validity and reasonableness. So, historically and in the present situation many hours in the production process are spent on editing the collected data. As mentioned above, the number of data sources that can be used for cross checking data is limited on a weekly basis. The result is in practice an information shortage in the very short run. There is not a complete set of updated primary information for each of the single variables in the BR at a given point in time. Instead, the information available must be used to make as good decisions as possible by the BR-team. From an incomplete set of information, and from years of

experience, the team needs to decide when hands-on editing is necessary and in some cases to contact the legal unit directly (by telephone or e-mail) in order to sort inconsistencies out. For instance, a large change in monthly gross wages for a legal unit may indicate an ongoing structural change that needs to be accounted for in the register as well as in the sample frames. A change of address may likewise indicate an ongoing change with implications for the structure of the enterprise and the compilation of its variables. Most changes in enterprises with more than 15 employees and enterprises with two or more local units result in a contact with the enterprise for verification before entered into the BR.

Some could argue that the supporting editing systems created for this situation also might risk to restrain the surge for new data sources, but the fact that new sources - especially commercial sources - has been added over the years indicates this is not the case.

Looking at the production in the longer run, for instance when producing yearly frames, the information shortage diminish. Then the creation of frames can rely to a larger degree on an increasing number of administrative registers. Here, time lags of several months is not so much of an issue. Examples of administrative registers are the yearly income statements, yearly accounts and standardized accounting statements.

Different opportunities for development

Main users of business statistics and national accounts press for access to statistics faster and with the same or improved level of quality. The increasing importance of the BR as the backbone for business statistics also place higher demands on streamlining the production of the register.

Trying to change or impose new regulations in order to improve statistics is hard. Not only need proposals to be treated, circulated for comments and ultimately passed by the national parliament, the general tendency on the view on data provision is likely to be the opposite - to lessen the burden of data provision on businesses and individuals.

Improving the BR by adding new sources or replace others is another option. This can be illustrated by the inclusion of commercial data providers. In one case, Statistics Sweden was actually data provider but chose to transfer the assignment to a commercial supplier instead. Since then the BR relies on deliveries from this particular supplier.

In a broader perspective, of course, the main suppliers of data remain the same today as three decades ago and the most important one is the Tax Agency. Therefore, when it comes to the legal framework and

access to data, the BR could hope for changes in the tax laws as much as in regulations concerning direct data provision from enterprises.

Over the last decade, one important change has been proposed in somewhat different shapes a couple of times: the monthly income statements. The proposals have all been voted down, until just now, and as far as can be perceived at this moment this administrative data will provide a great opportunity for developing the BR.

The point presented in this paper is however to turn to more effective ways to collect data by using web-based techniques. But getting access to a new set of administrative data does not contradict this movement to increased digitalization of the data collection process.

Digitalisation

One guiding principle for authorities launched by the Swedish government is to choose digital services for interactions with enterprises and individuals. This initiative also points out that basic information should be handed in only once to government agencies. As of now, it is under investigation to appoint one single authority to pursue the work on digitalization at national level. That is the case for instance in Denmark, Finland and Norway.

Instead, government agencies have voluntarily formed, joined and cooperated in different projects aiming to simplify data provision for individuals and enterprises. One such cooperation resulted in the presentation of a structure for data collection, data sharing and communication between businesses and authorities. Here, all compulsory data provisions to government agencies could be defined and carried out, using e-ID for log in.

For different reasons, the full realization of the structure has not been finalised. Some parts are in place, though. The most interesting from the BR-perspective is a first version of basic information – called Common information – for businesses. This information is what would be requested from businesses using different e-services supplied by those authorities that are connected to the infrastructure. The common information could be prefilled upon login to simplify for the businesses. Furthermore, the common information should be editable by the businesses. To add, withdraw or change information should be a built in feature of the portal.

Statistics Sweden together with the Tax Agency and the Company Registrations Office provide the data. Currently, information on local units for multi-establishment businesses is provided from the public BR. Businesses are presently able to view a limited set of basic information on its legal and local units and are also able to ask for changes of this information. The structure is implemented in the portal Verksamhet.se, which is used to register new businesses, get

general information and access services that are related to running a business in Sweden. The portal is a collaboration among several Swedish government agencies.

What still remains to implement is a way for all users of the portal to identify themselves via e-ID. Some legal units can already use the e-ID. However, the national initiative to implement e-ID for the use of e-services at government agencies has been delayed.

Direct reporting via web

Statistics Sweden is building a web-portal for direct reporting by all legal units. By the use of API, a business will be able to log in and access its information in the public BR. The technical structure for sending information to the Common information mentioned above is partly used here as well. No data will be shared between government agencies; this will only be a communication between respondents and the BR. The businesses will also be able to make changes to data which is sent to update the BR.

An advantage is the fact that all kinds of respondents will be able to log in, which is not yet the case in the portal with Common information. For the BR this is welcomed as most legal forms, including public sector, will be able to use the reporting via web. At Statistics Sweden an ongoing project regarding identification and authorization via a log in function is about to be implemented. It will use e-ID but also user-IDs and passwords. Thus there is a prerequisite for respondents to submit data to several surveys at Statistics Sweden as well as to utilize other e-services.

For the BR this solution will demand a higher degree of self-editing by the one supplying data for the business as data will be possible to hand in at any time. Compared to the present way to collect, edit and update information in the BR, this method will require a higher degree of fine tuning of the controls that will be built in. In order to get a smooth service for the respondent the portal must not be burdened by too many steps and checks before data can be sent in.

When respondents supply new och updated information to the BR, it must also be possible to view these updates. Presently, the idea is to make changes visible to the respondent the next day. Whether this period is too long or not can be discussed. Before changes of a larger nature are allowed to update the register there must be (we think) time for manual controls by the BR-team.

Anyhow, the use of direct reporting will most likely result in a change of effort regarding the editing process compared to today. It can be related to the time spent on editing of the yearly BR-survey. The surveyed legal units can respond during a period of three months, and although the survey is web-based with some built-in

controls, the data collection team assigned to edit the survey almost amount to six full-time persons. With the direct reporting responses will most likely be spread out over the year and not be concentrated to a specific period in the calendar. Whether this will affect the total amount of labour input for the process remains to be seen. Certainly, the way resources for maintenance is used will be forced to adapt to the changes in the editing process.

Web scraping

Web scraping as a means for retrieving information on legal units from the web is tested on an experimental basis at Statistics Sweden at the moment. An evaluation of the method is to be produced later this year. Therefore this section should not be thought of as a review of the method, but rather a presentation of a technique that could be of interest to the BR.

The work started in the Department for development of Processes and Methods. Tests is now ongoing to see whether web scraping can function as an alternative to current data collection for one specific enterprise based survey at Statistics Sweden. The ultimate goal is to reduce burden on respondents and to examine new ways to collect data. In these tests Statistics Sweden cooperates with a commercial actor involved in the field of business intelligence.

In the search of information, focus is on webpages, social media and e-trade. Combined with basic business information such as turnover, NACE and regional locality, these variables may be used as filters in a search operation.

The method is event-based which means, very simply, that if something new is published on the web this is detected and can be reported to the user. It is also possible to control different functions, like chats.

For the Swedish BR it would be valuable to be able to gather much of what a business presents on its web page. For instance the URL, but also addresses. For large businesses information on acquisitions and divestments are of interest. Of course it is of vital importance that the matching procedure is well functioning. Also, to value a piece of information correctly, the reference date or period is equally important.

The method is still under development and it will be exciting to take part of the evaluation later on. There may be shortcomings initially, but nonetheless we believe it is a method that will be a part of the data collection in the future.

Conclusion

In order to make the production of a BR more efficient, utilising digitalization to improve data collection is quite obvious. This has however been the case for a long time, but development has not gone as fast as predicted. The many small steps of development over a decade or so, by Statistics Sweden and other government agencies, can hopefully soon be realised. There are signs that this might be the case: businesses press to reduce response burden and to communicate more easily with government agencies, there is increased willingness to coordinate development on digitalisation at national level and solutions like e-ID are ready to be implemented.

For the Swedish BR, this will affect the methods used for data collection, but also the way we look at the editing process and ultimately how the resources are allocated in the production of the BR. Administrative data will likely be of better quality once reported and updated in a more standardized way to Statistics Sweden and other government agencies. This allows for new openings in how resources are allocated. The proportion of maintenance of the statistical BR in relation to the administrative BR can be increased, for example.