Business Registers as a Tool for Linking various Administrative Data Sources

Norbert Rainer
Statistics Austria

Paper presented at the

International Statistical Conference

Statistics – Investment in the Future 2

Prague, Czech Republic, 14 – 15 September 2009

Business registers as a tool for linking various administrative data sources

Norbert Rainer, Statistics Austria

Summary

In the past few years the role of the statistical business register has undergone a significant change. Being before mainly a database for performing statistical surveys with updating the register mostly from survey results, the business register now is a linkage system between administrative data sources and statistical requirements. The main sources for the updating of the business register today are administrative data. This has consequences on the structure of the register, the workflows, and the quality dimensions as well as on the use of the register which has increased consequently. It also opens new possibilities for the use and integration of administrative data sources into the statistical survey system.

In the paper, the Austrian developments and situation will be presented with a view to the general aspects of this new role of the business register serving as a linkage system to administrative databases. The problems encountered are of course dependent on the kind of the administrative data system, whether the administrative system has common identifiers, a rich list of variables and thus a broad coverage of data available, and similar aspects. In the Austrian case for instance, there is no common identifier so that matching tools are an important requirement.

1. Introduction

In the last few years we have observed some big changes in the development of official statistics. This was not only a result of an increased demand of statistical information, but it also concerned the production methods of official statistics. New statistical and survey methods, including electronic questionnaires and web-based data collection, have enforced the methodological tools available. The need for increased timeliness and, generally, a high data quality were further aspects that have driven the methodological developments.

However, the increasing demand of statistical data has also found its limits in the simultaneous increase of response burden (and also in the limited resources and budgets). In order to reduce the response burden various efforts have been undertaken, one of the most important is the multiple use of data surveyed and the use of already existing data in government files. However, this is not new for official statistics: already in the past a lot of statistical domains were fully based on administrative data (e.g. statistics on population movements). However, nowadays, the use of administrative data is not an option any longer, it is an obligation.

In the process of increasing the use of administrative files for business statistics purposes, the business register plays an important role. On the one hand, the business registers themselves are based on administrative registers; on the other hand, the business registers provide the link between the administrative input data and the statistical output data. Based on these links data from administrative files can be used to supplement or to replace survey variables or even whole surveys. This increases efficiency in the production of statistics, improves coherence of data and reduces response burden.

This paper aims to provide an overview of the current and potential future role of the statistical business registers in the system of business statistics with emphasis on the use of administrative data. The reference to the Austrian situation serves just for illustration.

Our conclusion is that the status of the business register with respect to co-ordination, coherence, efficiency and quality will increase. In the context of the necessary continuing redesign of the business statistics system, the business registers will have to play a central role.

2.   Functions of the business register

Figure 1 shows schematically the various functions of the statistical business registers. These functions are depicted around a triangle. On the left hand side of the triangle one finds the more traditional functions, the bottom side shows the functions that have lately emerged from the system of registers and the right hand side shows the recently growing functions, with the linkage function to administrative sources at the top. In the following subchapters all these functions will be described in brief.

Figure 1: Functions of the Business Registers
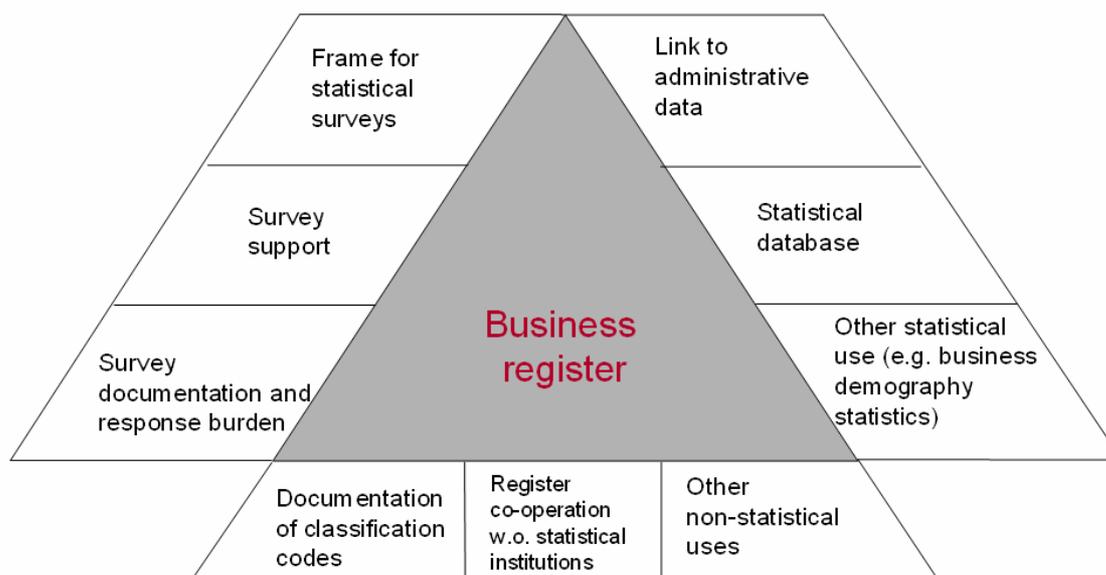
2.1   Frame for statistical surveys

Many years ago statistical business registers were just a list of addresses of legal units/enterprises, local units or other statistical units. The characteristics stored were the ID data of the unit(s), the activity code and perhaps some other stratification variables, such as employment size classes. Often the register did only cover one type of unit, so that even the relations within an enterprise to its local units or kind-of-activity units were not part of the system. The main function was to provide appropriate ID and address data for statistical surveys, which mostly were complete counts. The results of the economic surveys were an important basis for the updating of the register.

Later on, when sample surveys were the rule instead of complete ones, the business registers had to provide the frame for the sampling and the appropriate grossing-up factors. This meant that the business register had to increase its quality as only a part of the population was still checked by the surveys. The business register really served as the overall frame. Today in many countries no complete economic surveys are conducted any longer and the business registers have to use other data sources and methods for the replacement of information coming from surveys in order to update the register data. This holds true especially for the activity code and other classification codes of the units which are purely statistical codes that can only be attributed and updated if the appropriate information from the units is collected. Other information on the units was replaced by using administrative data.

This first of the various business register functions includes thus also an important co-ordinating and coherence role which is fulfilled if all the surveys addressed to enterprises and the other statistical units are based on the business register. This may in practice not always be the case, even within a national statistical institute (NSI). For example, some areas which were earlier not covered in the business register may still be surveyed using other survey frames. This may also be the case when some surveys are conducted by other institutions, such as the central bank, the environment agency,

or other government institutions. Whether these institutions can use the business register of the NSI will most likely depend on the national legal basis. However, it would be very preferable if all surveys used the same survey frame.

2.2   Survey support

The function of survey support is also a quite traditional one, but of great practical importance. On the one hand, the business registers store in addition to the postal addresses information on the units which are important contact data, such as telephone, fax number or e-mail address, name of the person responsible in the enterprise for the filling in of the various questionnaires. On the other hand, the business registers do also document when a questionnaire has been dispatched to the enterprise, when it came back, support sending reminders, or other information in connection with the conduction of a survey.

As many bigger enterprises have to report data for various surveys this function of the business register is also important from a coordination point of view within the statistical institute. Different surveys are conducted by different organisational units in the NSI and thus the contact data are available for all organisational units of the NSI.

2.3   Survey documentation and response burden

In addition to the survey support function, the business register is also the right basis for comprehensive survey documentation. Such survey documentation will record over time which units were selected for which surveys. This should be done preferably for all surveys, also for those not under direct responsibility of the NSI, and, furthermore, also irrespective whether the surveys were obligatory or voluntary. On that basis appropriate analysis of the distribution of survey obligations over the enterprises and the other statistical units can be performed and appropriate survey policies based upon. For instance, a policy of certain survey holidays can be established and with the help of the business register performed and monitored.

This documentation can also serve as the basis for response burden calculations and analysis. It can help to analyse and report the survey obligations over various variables, such as economic activity, size class, legal form, or number of surveys for which data have to be reported by the enterprises (for Austrian data see Rainer, 2004b, and Fröhlich, Rainer and Oschischnig, 2009). Based on such data, the results of a policy of reducing the response burden can be reported.

The survey documentation can also be enlarged by covering information on the response burden itself, such as the time needed to fulfil the survey obligation, or the costs of the time consumed in the enterprises. This can be used for the calculation and monitoring of total response burden over time, either for each single survey, for each economic branch, for each size class, or all in all for the grand total of response burden (for the Austrian methodology of the response burden barometer, see Rainer, 2004a and 2004c).

2.4   Documentation of classification codes

One of the most important stratification variables of the business register is the activity code of the enterprise and the other statistical units. The activity codes not only determine the structure of the industries and branches in the survey frame, also the coverage of many surveys is defined on the basis of the activity classification. Therefore, from the enterprises' point of view the activity code quite often determines their reporting obligations.

It is thus very important that the activity codes in the business register are of high quality. The determination of the activity code needs empirical and up-to-date information on the production structure of the enterprise and an indication on the relative importance of each of the activities performed. This could be data on employment or preferable data on the production value or turnover. In any case, the correct attribution of an activity code needs data reported from the enterprise, either integrated directly or indirectly into the regular economic surveys or as a special survey just dealing with the classification issue.

In the Austrian case the NSI has to notify the activity code to the enterprise. This is done when an enterprise is recorded in the business register for the first time, when the ID data are changed (name, address, legal form) and of course when the activity of the enterprise has changed. The enterprise should confirm the activity code attributed to it or, if it does not agree with the code, provide appropriate information on its production activities and profile. Based on this information the code might be changed and again sent to the enterprise for confirmation. This procedure is also documented in databases linked to the business register. For each enterprise one can observe the activity codes that an enterprise had over time. (For more information on the Austrian system and the efforts taken to ensure high quality of the activity codes see Karner and Rainer, 2007.)

2.5   Register co-operation with other statistical institutions

Depending on the structure of official statistics in a country the system might range from a completely centralised one to a highly de-centralised one. The more the system is de-centralised, the more it is important that even then the survey frame function of the business register can be fulfilled for as many surveys as possible. Again legal constraints are to be considered. Co-operation with other institutions of the statistical system might result in the need to document the exchange of data or to maintain additional IDs from the partner system(s).

A new form of register co-operation is currently under development between the NSIs of the member states and Eurostat in the case of the EuroGroup Register. This co-operation is based on the new EU Business Register Regulation set into force last year. The aim is to develop and maintain a register of the most important European enterprise groups. The co-operation includes the exchange as well as the checking of data on the European enterprise groups and their constituent legal units. This co-operation shall also comprise the European System of National Banks and their relevant data sources.

In Austria the system of official statistics is quite centralised. A few years ago a register co-operation with the Austrian National Bank has been developed. This includes currently mainly the delivery of certain register data to the Central Bank and the co-operation in the coding of the units in the financial activities (NACE) and the financial sector (institutional sectors of the ESA95).

2.6   Other non-statistical uses

Depending on the legal basis, data of the business register may also be used for non-statistical purposes. In some countries the register data are even publishable or are provided for administrative purposes. Whether this is the case or not depends on a number of circumstances, the legal basis is certainly the most important one. In other countries the statistical business register can only be used for statistical purposes.

In the Austrian case the NACE codes of the enterprises are forwarded to the social security institution and the tax register. Both data bases are also main data sources for the maintenance of the business register. The social security institution, for instance, elaborates the official employment statistics, and for reasons of coherence the employer enterprises in the social security register should have the same activity codes as in the business register.

2.7   Other statistical uses

In addition to the function as a statistical database (see below 2.8), the information of the business register may also be used for other statistical purposes. One recent example is business demography statistics. While data in the business register provide information on newly established enterprises or closed enterprises, these data might not be sufficient from the business demography point of view. According to the business demography concept, real births and deaths are only those events where no other enterprises are involved. The creation of a new legal unit is not a sufficient criterion for an enterprise birth. For an enterprise birth new factors of production, especially new jobs, have to be created. New enterprises created by break-ups, mergers, split-offs, take-overs or restructuring are not considered as newly born enterprises. Even if the business register cannot fulfil these requirements in total, the Eurostat Manual recommends to use the database of the business register as an important starting point for the derivation of business demography statistics (see Eurostat – OECD, 2007; for the Austrian situation see Haslinger and Rainer, 2009).

A second example is the new FATS statistics (inward FATS: data on foreign controlled enterprises resident in the country; outward FATS: data on foreign enterprises controlled by resident enterprises), where in the Austrian case administrative and survey data from the National Bank are to be linked to structural business statistics data which are compiled by the NSI. The business register provides the linkage between the two universes of units as well as certain characteristics for the FATS statistics, most important the NACE code.

2.8 Statistical database

As the business registers are mostly maintained by administrative sources, the coverage of the business register is very high with respect to activity and size classes. Depending on the sources available, the business register may cover not only enterprises, but also local units or other statistical units (new development: enterprise groups). Having up-to-date links to administrative sources, data on turnover and employment may also be available in a very timely frame. All this together transformed the business register from a survey frame system to a powerful statistical database.

The business register is a unique data base because of its coverage, because of its richness in units and central characteristics, because of its regional details, because of its timeliness, just to mention some of its main assets. Thus, efforts have increased to provide this information not only indirectly via the various surveys based on the business register, but also directly in form of regular business register statistics. It is further planned – not only in Austria - that in the next census round the survey on local units will be replaced by register statistics.

2.9 Link to administrative data

The maintenance of the business register is today mostly based on administrative sources. Normally, more than one source has to be used. In many countries the most important sources are tax registers, social security registers, company registers, and registers of the chamber of commerce or similar institutions. In addition, a lot of other specific data sources are used, in Austria for instance, the register of non-profit institutions, the school and the hospital register, and others.

Whatever sources are used, the IDs of the specific source are stored in the business register as a link to the respective unit in the register. This link allows to update the register data whenever the source data are changed as well as to shift specific variables of the respective source data to the business register unit. So, for example, the link to the social security register allows attributing each month the numbers of employees from the social security database to the business register for each single employer enterprise. The same is true for turnover data from the tax register. This enables to elaborate very timely data on employment which can build the basis for appropriate employment statistics, or replace employment variables in certain surveys.

The link to the administrative registers is not only used for the provision of the variables to be recorded in the business register (such as employment data), but for all other data in the administrative databases that are used in certain statistical domains. For example, data on wages and salaries of the social security databases are used for the labour cost statistics, data on income from the taxation database for the income statistics, etc. The links of the IDs in the business register provide the exact matching of the units in the administrative database and the statistical units in the business register.

3. Use of administrative data

Administrative data today are the most important source for the maintenance of the statistical business registers. Their use generally improves all aspects of quality of the business register, especially coverage and timeliness, and – as shortly described above - opens new possibilities for supplementing or replacing statistical surveys. Both types of uses reduce the response burden considerably. However, the administrative data sources partly follow different definitions and rules than those needed or desired for statistical purposes. This makes it sometimes difficult to use these sources in a straightforward way without any additional efforts and data editing procedures. Furthermore, the availability of administrative data is not always sufficient for business register purposes and the data are of different quality. Administrative databases are also at risk to be changed

suddenly. It is therefore important to keep alive a good co-operation between the NSI and the holders of administrative data, as well as to undertake common efforts in order to achieve improvements of the data quality.

Normally, more than one administrative data source has to be used for the business register maintenance, either because the coverage of the source is not sufficient or because the variables needed are only available in various sources. In the Austrian case four main administrative sources are used:

- Register of companies (includes - with exceptions - only corporations)
- Tax register (is the most complete register, also used for turnover data)
- Social security register (includes only employer enterprises, also used for employment data)
- Register of Austrian Federal Economic Chamber (includes only units in needing a business licence; does not cover free-lancers; provides data on local units).

An enterprise may thus be covered by one of these registers only, or by two, or by three or by all four of them. This is per se no problem; however, the issue is that these registers are maintained independently and the notation of the name, address and other ID variables might differ considerably. Even that would be no problem if a common identifier would exist. However, in the Austrian case there is no single (or common) identifier for enterprises/legal units. Therefore, specific procedures of string matching methods have to be applied to determine the links between these sources (see Haslinger, 2004; Denk, Hackl and Rainer, 2005).

One enterprise/legal unit can be linked to more than one unit in a specific administrative database, because the administrative register is run decentralised (regionalised) and each of these registers covers only a part of the whole enterprise. Additionally, the profile data of the various register units that would belong to the same enterprise do not necessarily coincide. The task of matching is also hampered by the fact that the administrative data are updated at different points in time.

A further problem is that the continuity rules in the administrative registers differ from the statistical rules. Changes of legal form disrupt continuity in most administrative registers, however, to a different extent. For most of the administrative registers a change in legal form means a different treatment of that unit and thus a new unit with a new identifier is created. However, for statistical purposes a mere change in legal form does not disrupt continuity. As already mentioned above in the context of business demography data, a new identifier code in one of the administrative registers does not necessarily mean a new statistical unit. However, information on predecessors and successors are normally not available in most of the administrative databases which clearly increases matching problems.

These few examples should show that the use of administrative data is connected to various conceptual and data problems (for the Austrian case see Rainer, 2008) and that many efforts have to be taken to overcome these deficiencies as good as possible. Concerning the experience in the use and treatment of administrative data for the purpose of the business register in other countries see for example: Anderssen, 2008 and Egmose, 2000 for Denmark; Statistics Finland, 2007 and Viitaharju, 2007 for Finland; Ritzen, 2001 for the Netherlands; Romanov, 2002 for Sweden.

On the other hand, there is no other way as to use administrative data to the largest extent possible, in general and especially in case of the business register. It is furthermore much more efficient to establish the links to the administrative data and to solve the related data problems in the context of the business register than if this has to be done in the various statistical domains separately. The co-ordinating role of the business register is also valid here.

4. Resume

The great importance of the statistical business registers in a system of official statistics is certainly generally acknowledged. The business registers (similarly the other statistical registers in other statistical domains) fulfil basic functions for the production process of statistics and the quality of the statistical outputs. This importance has increased as new functions have emerged, such as the use and linkage to administrative data. Furthermore, the business register has to be embedded in a

system of statistical registers, in particular to the population register and to the buildings and dwellings register, with the appropriate linkage system between the units in these registers.

In a recent communication of the European Commission on the production method of EU statistics (European Commission, 2009), the current production method is described as the stovepipe model: "For each domain, the whole production process from survey design over data collection and processing to dissemination takes place independently from other domains, and each has its own data suppliers and user groups." This production method is to some extent still prevailing at the national as well as at the European level. The Commission sees at least three main disadvantages of the stovepipe model: it imposes an unnecessarily heavy burden on respondents (respondents are asked for the same information more than once), the model is not fit for statistical domains with multiple dimensions (e.g. globalisation), and thirdly, the production is highly inefficient and costly (because of missing standardisation and co-operation).

In order to overcome these deficiencies, the production methods should be re-oriented towards an integrated system (data warehouse approach) together with the appropriate use and re-use of administrative data for statistical purposes. For this re-orientation of the production method of statistics, the business registers will certainly play an important role.

References

1.   Andersen, S. Sch.: *Adaptation of the statistical business register to changing external conditions*, 21st International Roundtable on Business Survey Frames (Wiesbaden Group on Business Registers), Paris, 24-27 November 2008.

2.   Denk, M., Fröschl, K.A., Hackl, P. and Rainer, N. (Guest Editors): *Data Integration and Record Matching, Proceedings of the DIECOFIS Workshop in Vienna, 13-14 November 2003*, Austrian Journal of Statistics, Vol. 33, Number 1&2, 2004.

3.   Denk, M., Hackl, P. and Rainer, N.: *String Matching Techniques: an Empirical Assessment Based on Statistics Austria's Business Register*, Austrian Journal of Statistics, Vol. 34, Number 3, 2005.

4.   Egmose, S.: *Administrative units, the BR and the business statistics*, 14th International Roundtable on Business Survey Frames, Auckland, 29 October – 3 November 2000.

5.   European Commission: *Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade*, COM (2009) 404 final.

6.   Eurostat: Business Register Recommendation Manual, 2003 Edition, Luxembourg, http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-BG-03-001-   -N-EN.pdf.

7.   Eurostat – OECD: *Manual on Business Demography Statistics*, 2007 Edition, Luxembourg, http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf.

8.   Federal Statistical Office of Germany: German *Business Register re-engineering: Heading for a tool to acquire, combine and connect statistical data sources*, Joint ECE/Eurostat/OECD Meeting of Experts on Business Register, Geneva, 18-19 June 2007.

9.   Fröhlich, M., Rainer, N. and Oschischnig, U.: *Meldepflichten und Belastung der Wirtschaft durch Erhebungen der Statistik Austria 2001 – 2008*, Ergebnisse des Belastungsbarometers, Statistische Nachrichten, Heft 6, 2009.

10.   Haslinger, A.: *Data Matching for the Maintenance of the Business Register of Statistics Austria*, Austrian Journal for Statistics, Volume 33, Number 1&2, 2004.

11.   Haslinger, A. and Rainer, N.: *Calculating Business Demography Statistics based on Administrative Data*, European Establishment Statistics Workshop, Stockholm, 7-9 September 2009.

12.   Herschap, N.: Co-ordinating role of the Business Register in Economic Statistics, 18th International Roundtable on Business survey Frames, Beijing, 18-22 October 2004.

13.   Karner, Th. and Rainer, N.: *Measuring and improving the NACE coding in the business register*, 20th International Roundtable on Business Survey Frames (Wiesbaden Group on Business Registers), Wiesbaden, 21-26 October 2007.

14.   Rainer, N.: *Measuring response burden: The response burden barometer of Statistics Austria*, European Conference on quality and Methodology in official Statistics, Mainz, 24-26 May 2004, 2004a.

15.   Rainer, N.: *Meldepflichten bei den wirtschaftsstatistischen Erhebungen der Statistik Austria im Jahr 2003*, Statistische Nachrichten, Heft 5/2004, 2004b.

16.   Rainer, N.: *Measuring response burden in a system of Total Quality Management*, International conference "Statistics – Investment in the Future", Prague, 6-7 September 2004, 2004c.

17.   Rainer, N.: *Quality of administrative data – a challenge for the maintenance of the statistical business register*; ECE Work session on Statistical Data Editing, Wien 21-23 April 2008.

18.   *Regulation (EC) No 177/2008 on the European Parliament and the Council of 20 February 2008 establishing a common framework for business registers for statistical purposes and repealing Council Regulation (EEC) No 2186/93*, OJ No L 61 of 5.3.2008.

20.   Ritzen, J.: *Managing relationships with Administrative Agencies, 15th International Roundtable on Business Survey Frames*, Washington, 22-26 October 2001.

21.   Romanov, H.: *The Swedish Business Register in relation to administrative sources and the Swedish Register System*, 16th International Roundtable on Business Survey Frames, Lisbon, 21-25 October 2002.

22.   Statistics Austria: Standard-Dokumentation Unternehmensregister, 2009, http://www.statistik.at/web_de/dokumentationen/Unternehmenarbeitsstaetten/index.html.

23.   Statistics Finland: *Business Register and Administrative Data – Ground for effective production of economic statistics*, Joint ECE/Eurostat/OECD Meeting of Experts on Business Register, Geneva, 18-19 June 2007.

24.   Viitaharju, T.: *Managing Administrative Sources, the Business Register as a Link between Economic and social Statistics*, 16th International Roundtable on Business Survey Frames, Lisbon, 21-25 October 2002.

25.   Viitaharju, T.: *Business register as basic tool for the production of statistics*, International Roundtable on Business Survey Frames (Wiesbaden Group on Business Registers), Wiesbaden, 21-26 October 2007.