

Contents

- 1. Introduction**
- 2. Categories of data source**
- 3. Common issues**
- 4. Administrative data source**
- 5. Advantages of using Administrative data**
- 6. Disadvantages of using Administrative data**
- 7. Quality of Administrative data**
- 8. Legal issues and access to administrative sources**
- 9. Using administrative sources in practice**
- 10. Converting to statistical units**
- 11. Quality improvements and addition of specific stratification variables**
- 12. Statistical Sources**
- 13. Combining Administrative and Statistical sources**
- 14. Links to other types of registers**
- 15. Other and New types of data sources**

Chapter 6: Data Sources for the Statistical Business Register

1. Introduction

This chapter provides guidance on sources of data that can be used to construct a business register. The main focus will be on how to use administrative data to construct a business register, however the chapter will cover other sources of data such as statistical sources. The chapter will define administrative data and will describe the administrative sources that are used by many countries. The advantages and disadvantage of these sources will be considered. The chapter will consider legal aspects of administrative data, as well as thresholds etc. Finally the chapter will look at other possible sources of data for the business register. The chapter will not provide guidance on how to build databases based on administrative sources.

It is recommended that countries establish statistical business registers which are mainly fed by administrative sources to ensure stability and high coverage. This is however not always feasible for a number of reasons and how to deal with this situation is important to many countries.

Different sources may be utilized in setting up a statistical business register. If the coverage of the statistical business register is to be as representative as possible, it should contain current information on its constituents. This means data sources should be able to maintain the register over time and take note of the changes in enterprise dynamics. Unless the statistical business register is maintained on a regular basis, it will quickly lose its value by becoming dated and ceasing to adequately reflect activity in the real world. Described below are some of the sources utilized for creating and maintaining a statistical register of business units.

The structure of the chapter is first common issues relating to any kind of source will be presented after which each category of source will be dealt with and the different sources will be discussed. Examples will be provided where relevant.

2. Categories of data source

Data sources depend on the specific country situations, such as the availability of certain administrative data and registers, as well as the statistical system. However, the most common sources can be described and grouped into three categories:

1. Administrative sources
2. Statistical sources (incl. feedback from surveys)
3. Other sources (e.g. private data suppliers)

2.1 Administrative sources Administrative sources are today usually the most important source for business registers. Therefore, providing guidance on how to establish and maintain good cooperation with the owners of administrative registers is important.

A Eurostat Task Force in 1997 defined administrative data as:

“Administrative sources are sources containing information that is not primarily collected for statistical purposes”.

This definition allowed for a very wide interpretation of administrative data. A more traditional definition relates data collected by government bodies for the purpose of administering taxes and public pension funds. But administrative data might also be available from private companies, e.g. telephone companies, water or power supply etc.

Administrative sources also include associated registers which can be defined as .????.. . These kind of registers will not be directly used as source for updating the Statistical Business Register, but as quality check, see section 11.2.

The current ESSnet on Administrative data, a Eurostat funded project group, defines administrative data as:

The data derived from an administrative source, before any processing or validation by the NSIs.

Eurostat collects information on the administrative and other data sources used for statistical business registers in the annual business registers questionnaire. This questionnaire is also used by the UN/ECE so covers a wide range of countries including the OECD countries. The most commonly used sources relate to taxation systems such as value added tax (VAT) and personal income tax or to compulsory business registration systems, often administered by chambers of commerce

Statistical sources are output from the statistical process primarily done within the national statistical institute itself but in principle it could come from other producers of statistics. The statistical sources have most form of feedback from different statistical domains.

2.2 Other sources

Other sources are of commercial kind and could be collected by companies for commercial purposes. For example companies such as Dun & Bradstreet or Bureau van Dijk. Although these sources are used in SBRs for creating Global Enterprise Groups, the chapter will not include these in detail

3. Common issues

3.1 Cooperation

Cooperation does not limit itself to the administrative sources although that cooperation often proves to be the most complex and challenging.

The purpose of cooperation from the point of the statistical business register is to ensure continuity, to reach conceptual approximation and to ensure easy matching of different sources, preferable through a common identifier; in short to promote the interest of the statistical business register and in the end business statistics.

The purpose is the same whether the SBR are dealing with an administrative, statistical or commercial source.

The cooperation can take on many forms but it is recommended that it involves regular face-to-face contact and mutual visits. The contact should also involve both management level and floor staff working day-to-day with the data.

Through cooperation it is possible to be proactive regarding changes to the register to either limit unfortunate changes or at least to have a head start on changes that might otherwise disrupt the statistical production. The cooperation should also be used to promote the statistical concepts in the administrative register, e.g. code of activity or definitions of business units.

A common administrative identifier eases the work with business statistics and if one does not exist, the national statistical institute should actively promote one among the data providers and even bring the providers together to show the possibilities in a common identifier. The goal of linking different data is not only an argument for statistics but also for administration. As long as a common administrative identifier does not exist, the NSI should construct one itself, at least for the Legal Unit (LeU). It is important to be aware of the difference between an administrative unit and the LeU, e.g. the same LeU in some countries can have several administrative units in the same administrative register.

3.2 Identify useful sources

The first step is to identify the most useful source(s), evaluated by information and costs. It is important to acquire all available information, e.g. what definitions are used, what is the coverage, what updating methods are used and what is the updating frequency, what is the time lag and how often will it be possible to get information from the source.

3.3 Using external sources in the SBR system

This chapter is not a technical chapter laying out the good design of a statistical business register. However, when talking of sources a few important guidelines as to the design of the SBR should be mentioned.

It is strongly recommended to have one part of the register containing the external sources with their original content only extended with history keeping track of changes in the deliveries from the external sources, i.e. if a unit changes address in an administrative register the old address is stored with a termination date and the new address with a commencement date and open ended termination date.

Then through a set of conversion rules and procedures, the statistical part of the register is created. This is where any changes for statistical purposes is made.

3.4 Single Administrative Register

Some countries have developed a single administrative register, with multiple functions, including that of the statistical business register. This has many advantages, in that the statistical data requirement is within the core of the tax and administrative system.

One excellent example is the Norwegian Business Register (Brønnøysundregistrene), but also other Nordic countries and **Austria** have similar systems. In Annex **X** the history about the development of the Danish system is described.

3.5 Linking and matching

The SBR will be more resilient with lower risk, if multiple data sources are available. Some sources such as VAT, will have extensive coverage of the economy, whereas others may only cover one sector e.g. financial. So there are two linking/matching challenges; firstly where more than one source has a similar all economy coverage, and secondly where multiple sources only cover limited sectors, such as financial, and public administration. The first matching challenge is avoiding duplication or omissions in the whole process, whereas with the second approach the main concern is the sectoral boundary between each source.

If there is no common identification code, a probabilistic approach to record linking can be adopted by identifying correspondences or similarities in the name and address or other characteristics (e.g. legal form, economic activity code etc.). This process is generally referred to as matching. However, this approach makes it possible to link units in error (sometimes known as false matches).

3.5.2 Transforming administrative units to statistical units

The next step is to identify the link between units in the administrative source and the corresponding units in the statistical business register. This is relatively easy if there is a common reference number for businesses, as long as there are no errors in this number in either the administrative source or the register.

3.5.3 Quality

Quality checks of the results of the matching should be carried out. These can take the form of checks of auxiliary variables, e.g. economic activity codes, size or legal form. If these variables are consistent, it is more likely that the match is correct. If not, further clerical checks may be needed, particularly where larger units are concerned. But if administrative statistical information such as turnover and import/export are available from different administrative sources (as well as from statistical sources) these could also be used in the checking procedure.

It is also useful to periodically check the non-matched units and attempt to establish further links or to determine why they do not match. If the non-matched units are genuine, they may be due to timing or scope differences between the administrative source and the statistical business register. This could lead to find sources to get information on starts and deceases.

In some cases it is possible always to use the administrative unit as a statistical Local Unit (LU) or even a statistical Local Kind of Activity Unit (LKAU)

4. Administrative data sources

The definition used in this chapter on administrative sources is stated in section 2. Some administrative sources used for statistical business registers are those mentioned below. Besides information on name, address etc. these registers also contain information which is useful for statistical purposes and which can indicate whether or not the administrative unit is active or not. In annex X an overview over the most possible variables to get from the different sources and which have relevance to the SBR is shown.

Administrative Business Register records (compulsory administrative basis register for registration) – if available and if this is the case this basis register gives all basis information of ID-number, name, address and other contact information. This register could be run by the Tax Authorities, Chambers of Commerce or another public authority.

Tax records (e.g. those relating to VAT or employee income tax) may be useful to obtain and update register data on economic activities, turnover, and status of activity.

Company/trade, chambers of commerce registers can provide details of compulsory registration such as information on economic activity, legal form and birth and death of a unit.

Social Security register on companies employing paid staff, contributing to this social contributions for employees could provide identifying characteristics, and economic/stratification characteristics (legal form, employment).

Labour and Employment register for fiscal purposes can provide additional economic and social information about employees.

Industry associations may also be used as a primary and very up to date source in creating and updating the business register.

Agricultural administrative registers

Telephone directories or special listings prepared by telephone companies could provide also useful information to the business register.

Water supply and electric associations etc.

Central Banks have often information from supervisory authority for the financial sector and on units engaged in foreign direct investments flows.

Published Business Accounts: Countries are encouraged to explore the possibility of automatically extracting data from the internal financial or management accounting systems of businesses (e.g. using XBRL). Data from financial reports are in particular valuable for Business Registers as they contain information on shareholders and subsidiaries that is essential for building up enterprise groups.

5. Advantages of Administrative Data

5.1 Quality

Administrative sources often give complete or almost complete coverage of the target population, whereas sample surveys often only directly cover a relatively small proportion. The use of administrative sources therefore eliminates survey errors, removes (or significantly reduces) non-response and provides more accurate and detailed estimates for various sub-populations, e.g. businesses in small geographic areas. Coverage is often of

particular interest from the viewpoint of Statistical Business Registers, given that such registers should cover all economic active units.

The main advantage of administrative sources is the level of coverage, which will either be complete, or if not complete well defined, so that it is possible to assess the administrative data coverage against the target population.

5.2 Costs

Another main advantage administrative sources offer over survey data is comparative lower cost. Surveys are very expensive, particularly if they are conducted as censuses or involve the use of personal interviews. Administrative sources are often available free of charge, or for a small marginal cost of extraction, particularly if they originate from the public sector. Even if there is a charge, it is often still cheaper to use administrative data than collect the same information via other methods. Fewer staff are usually needed to process the data and there is no need for response chasing. The size and scope of statistical business registers makes it very difficult and expensive to satisfactorily populate and maintain them solely by economic census data.

5.3 Response burden

Using data from administrative sources also helps to reduce the response burden on businesses. Whilst businesses usually understand the reasons for supplying data for registration and taxation purposes, even if they do not like doing so, they may see statistical data requests as an extra, less necessary, burden. If they have already provided details to other government departments, they may become annoyed at receiving similar requests from the national statistical institute. An associated advantage is that the use of administrative data may, in some cases, allow statistics to be produced more frequently with no extra cost to businesses.

5.4 Quality

The use of administrative sources may increase the quality of business registers by allowing access to more up-to-date information concerning certain variables, e.g.

- Name and address;
- The opening and closing of units and the dates of these events;
- The economic activity code (at enterprise group, subgroup, enterprise and/or local unit level);
- Information on unit locations (at local unit level);
- Size data (number of employees or turnover).

5.5 Timeliness. As well as improving the timeliness of variables, the use of data from administrative sources can, in some cases, improve the timeliness of statistics that are derived from business register data either directly or via samples. This is because statistical surveys generally take time to plan, to design and pilot forms, to analyse the population and optimise the sample, etc. This is particularly the case for annual or ad-hoc data collections. Therefore access to a suitable administrative source via the business register can be a more efficient solution. It should be noted, however, that there are also likely to be cases where the use of administrative sources leads to a reduction in timeliness, particularly regarding short-term indicators, if the administrative data has long time lags. For example, some business tax returns may not be required until several months after the fiscal reporting period.

6. Disadvantages of using administrative sources

Although there are many good reasons for using administrative sources, there are also a number of common problems associated with their use. The following paragraphs outline some of these problems and propose methods to solve them, or at least to minimise their impact on statistical business registers.

6.1 Definitions of administrative units and variables differ from statistical definitions

One major problem often encountered when using administrative sources is that the units used in those sources do not correspond directly to the definition of the required statistical units. As well as differences in the definitions of units, there are also likely to be differences in the definitions of variables between administrative and statistical systems. The data in administrative sources has generally been collected for a specific administrative purpose and the needs and priorities relating to that purpose are likely to be different to those of the statistical system. For example, turnover for value added tax (VAT) purposes may not include turnover related to the sales of VAT exempt goods and services, whereas the statistical system is likely to require total turnover.

Similarly, the classification systems used within administrative sources may be different to those used in the statistical world, for instance a register within a food safety authority might classify a department store by its food department. Even if they are the same, they may be applied differently depending on the primary purpose of the administrative source. Where classification systems are different, it is usually necessary to construct conversion matrices to map the codes in the administrative classification onto those required for statistical business registers. Such mappings may be one-to-one, one-to-many, or many-to-many. In the latter cases, some sort of probabilistic allocation is often required. This should result in accurate coding at the aggregate level, but not necessarily at the level of individual units.

6.2 Timeliness

Another common problem encountered when using data from administrative sources relates to timeliness. Data may either not be available in time to meet statistical needs, or it may relate to a period which does not coincide with that required for statistical purposes, e.g. a tax year may not coincide with the calendar year required for structural business statistics. There will generally be some sort of lag between an event happening in the real world and it being recorded by an administrative source. This is then followed by a further lag before the data is made available to the statistical business register. Lags relating to births and deaths of enterprises are a major source of register coverage errors. If these lags can be measured, allowance can be made for them in any statistics based on register data.

6.3 Possibilities to meet disadvantages ????

Public sector administrative sources are generally set up for the purposes of collecting taxes or monitoring government policies. This means that they are susceptible to political changes. If a policy changes, administrative sources may be affected in terms of coverage, definitions, thresholds, etc. or possibly even abolished completely. Such changes may happen suddenly, with little warning, particularly following a change of government. Reliance on a particular source always carries a certain degree of risk. These risks can be managed to some extent by legal or contractual provisions, regular contact with those responsible for the administrative source to try to obtain early warning of possible changes and by drawing up contingency plans.

If data from several administrative sources is used, it is likely that the manager of the statistical business register will be faced with problems matching the data. Matching is relatively easy if there is some form of common identification number, but, if not, it usually has to be based on variables such as name, address and NACE code. In such cases, it is likely that there will be a certain proportion of false matches and false non-matches, giving rise to a need for the clerical investigation of possible matches. When there is a lack of a common identifier, the national statistical institute should use any opportunity to promote one and show the advantages of linking data in terms of more accurate and less costly statistics – not to mention administrative advantages.

Another problem where multiple sources are used concerns consistency between the sources. Data from one source may appear to contradict that from another source. This may be due to different definitions, classifications or differences in timing or simply to an error in one source. To resolve such conflicts it is necessary to establish priority rules, by deciding which source is most reliable for a particular variable. Once a priority order of sources has been determined for a variable it should then be possible to ensure that data from a high priority source is not overwritten from a lower priority source. This process is made much easier if source codes and dates are stored alongside the main register variables.

7. Quality and administrative sources

7.1 Create knowledge of the administrative sources

Although administrative data has many advantages, it is important to invest time in understanding and monitoring the quality of administrative sources. It can be useful to compare sources in terms of coverage and accuracy of the variables. This can be aided by a quality survey to determine the correct values of certain variables. The closeness of administrative units and variables to the units and variables required for statistical business registers can be an important factor in determining the quality of an administrative source.

Perhaps the best way to assess the quality of an administrative source is, however, to build up a thorough knowledge of that source, including the primary purpose of the source and the way the data is collected and processed. Thorough understanding of a source will allow a more accurate assessment of strengths and weaknesses. To help develop and document this knowledge, particularly for the benefit of future register staff, it is useful to develop some form of template to record information on contacts, units, variables, quality and formats. The information in the templates should be agreed with those responsible for the administrative source and should be reviewed regularly.

7.2 Quality indicators

Regular monitoring of quality is important, and it is recommended that a range of regular quality indicators are developed. Some examples of quality indicators are:

Random sample to look into activity codes, addresses etc.

Following the development in statistical variables within groups of activity codes to become aware of changes in administrative rules which can influence the activity code registered in the administrative registers.

Look carefully at changes in activity code for big companies.

MORE

8. Legal issues and access to administrative sources

8.1 Legal act

The first consideration is for the statistical office to get a legal gateway to the country's relevant administrative data. This will be critical to the development and maintenance of the business register. There are two aspects to gaining access to administrative sources; the first concerns the legal framework, whereas the second covers the practical issues surrounding the transfer of data. The legal framework will vary from country to country, but the preferred approach is to have an automatic right of access to administrative data enshrined in a general statistical act.

8.2 Actions if legal act is missing

In many countries such an automatic right does not exist and there might even be confidentiality legislation regarding the administrative registers actively preventing an exchange of data, even for statistical purposes. This can prove to be the largest obstacle to the establishment of the best possible statistical business register and not one solved quickly. The actual approach to solve the problem will be specific to each country, but the basic arguments in favour of access to administrative registers for statistical purposes is the same: cost efficiency, quality of statistics and reduction of burden to businesses.

It is recommended that some sort of formal agreement is obtained with administrative data suppliers, such as a legally binding contract with private sector suppliers, or a 'service level agreement' or 'memorandum of understanding' within the public sector. These agreements should describe the rights and responsibilities of both parties, data confidentiality constraints, quality standards, frequency and format of data transfer and procedures to follow in case of disputes.

8.3 Building Relationships with Administrative data suppliers In addition to formal arrangements, it is also strongly recommended to build a good working relationship with administrative data suppliers. This can be achieved through regular contact, preferably at least partly face-to-face. It is usually worth devoting some time to visiting suppliers to gain a greater knowledge of their work, systems and constraints. This leads to a better appreciation of quality issues and can help to build goodwill and mutual understanding, which in turn helps to identify and resolve problems as they arise and before they escalate.

Formal arrangements can be supported by Service level agreements, which provide written agreements on delivery flows, confidentiality, time to respond to questions about the data, and technical delivery details for example.

A main prerequisite for using administrative data is the access to it by the national statistical authority. The government and the administration might favour the sharing of data in some countries. In other countries, however, there may be reluctance at the thought of data being shared throughout the government. It is very difficult to reduce such concerns, but possible approaches could include the publication of clear limits and rules regarding the use of data, ensuring that businesses understand that sensitive data will not be fed back to other parts of government (particularly tax agencies). This is in line with the United Nations 'Fundamental Principles of Official Statistics', where Principle 5 (*"Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents."*) encourages the use of administrative data and taken in conjunction with Principle 6 (*"Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used*

exclusively for statistical purposes.”), establishes the principle of the one-way flow of data. The publication of analyses of the costs and benefits, both to government and to businesses, of the use of different sources may also be beneficial.

9. Using administrative sources in practice

The following sections have to be seen as a kind of steps in creating and maintaining a Statistical Business Register (SBR). How exactly the data model should be established is depending on the possible data access, but some basic circumstances are recommended.

9.1 Keeping administrative and statistical information separate

It is suggested to keep a separate set of data on the administrative units received from the administrative systems and not mix this information with the statistical units and data concerning these statistical units. Experience shows that it is much easier to discuss odd situations either with the business itself (which know what information has been given to the administrative source) or with the source if that should be necessary, giving concrete examples.

Which data should be a part of the SBR or be available in another way (for at least SBR staff), for instance directly on line from the administrative source or from a Data Warehouse, depends on the concrete situation.

9.2 Establish a unique identifier

The establishment of a unique identifier is essential for the accurate maintenance of the SBR and should be created for the statistical units. For administrative data where multiple data sources are used with multiple identifiers significant resources are required to match and create unique enterprises from these sources. Ideally the statistical agency would have influence over the development of administrative data, which would be used to develop a unique business number in the country. However, in practice this is not likely to be easily achievable, although there have been occasions where this has happened.

Multiple business numbers lead to added complexity in business register systems, and introduce risks such as duplication and omission of businesses in the population.

The administrative units – if they are not equal to a legal unit (LeU) – have first of all to be linked to the LeU and if possible to an administrative local unit (or if an administrative LKAU exists to this if possible).

But the most complex situation is the situation where a LeU might use several administrative identification numbers for registration in the same administrative register. In some cases a LeU is using several administrative identification numbers, but after some time it either closes some of these or asks for more. It is important to handle these situations. Usually, it will be possible to identify the LeU – which is in the end responsible to the public (whom you should communicate with if an account has to be paid, whom is responsible in case of an accident etc.) – and it will also be possible to pick out the ‘main’ unit in the administrative register to be the one which – as long as it seems correct – is the one used to feed information about the LeU. These actions are pointed out because it is important to see the difference between LeUs and LeUs owning other LeUs in a concern-structure, because one LeU is not responsible for other LeUs even though they are part of the same concern. So, first of all it is important to create the administrative LeU if it is not available from administrative registers, and then link the administrative units to this LeU.

Eventually some administrative sources identify local units or some kind of administrative LKAU. If this is the case it is important to be aware of whether these units are defined in the same way in all administrative sources or the only possibility is to create an administrative local unit (LKAU) for each source.

9.3 Sector specific sources

Some trade association type sources can be useful for certain sectors of the economy, but care needs to be taken to ensure that their membership is complete or gaps are understood. Also there may be administrative sources for certain sectors of the economy, for example charities might be regulated and could be a source of these, or schools could be identified from the Education Ministry. Using multiple sources such as the examples listed above will lead to challenges of maintenance and management. In some cases the source will be used to ensure that the administrative data is correct, in other cases the source will be a part of updating the statistical units. It is important that issues of duplication and omission are considered. However such sources can be good quality and worthwhile pursuing.

9.4 Thresholds

Thresholds such as a turnover limit (which means that a business does not have to register if the turnover is below this limit) have an impact on the coverage of the register. In many countries these thresholds do not have any real influence on the statistical figures. If this is the case it is not worth to bother about those businesses. But if a lot businesses are below the threshold supplementary sources are needed.

9.5 Changes to administrative sources

Changes to administrative sources such as changes in thresholds, definitions or computer systems, can all lead to discontinuity in the data supply. A set of key steps to managing this type of change, is to firstly work very closely with the administrative data supplier to gain a full understanding of the change. This understanding can then be used to assess the impact on the SBR. If possible, it would be good to process the new administrative data in a separate copy of the SBR database, so that the impact can be directly measured. Often there are unexpected consequences, when making changes to administrative data, when complex systems are in place. Once a good assessment of the impact has been made, the next step is to actively work with the SBR users to assess the impact on their outputs.

Two possible solutions exist. Firstly, to make a transformation in the SBR to minimise the discontinuity, or secondly, to let the users deal with the discontinuity through their own systems. The method used is clearly linked to the nature of the change. For example, a change the classification systems in administrative data can be handled by transforming data in the SBR, whereas an expansion of the register frame population caused by improved coverage may be handled by the users.

Changes in tax thresholds are common and often just reflect the general inflationary environment in the country. In which cases these may not have any real impact on coverage. Occasionally more radical changes in thresholds will take place, reflecting the political environment. These type of changes will have an impact on register coverage, particularly if a single sources is used. It will be possible to estimate the impact on the population by comparing business birth trends before and after the change, for example, but a permanent change in the population could result.

10. Identifying the statistical units

In this section the focus is on identifying the Enterprise and Local Unit/Local Kind of Activity Unit (Establishment).

10.1 Identifying the Enterprise

In most countries with administrative registers, a legal unit (LeU) of some sort is often defined or could be defined in the statistical office, e.g. above.

In many situations the LeU will be, if not perfect, then a good approximation to the enterprise. However, as is already mentioned in chapter X, it is necessary a least for the bigger businesses, to create an Enterprise in accordance to the definition of this unit. Whether or not this follow the definition of the smallest combination of LeUs or a profiling technique top-down using segments the ‘transformation’ uses some rules. These rules and processes should be put in here between the administrative units and the statistical units so it is clear how the administrative units contribute to the updating of the statistical units; which information overrules another and when did it happened. See section 10.4.

10.2 Identifying the Local Unit/Local Kind of Activity Unit (Establishment)

Local units are an essential component of the SBR, which enable regional economic analysis to be conducted. Many of the usual administrative sources, such as VAT are only available for the LeU as a whole or a part of this, but not necessary at local level, so they cannot be used to populate local units. Therefore creating local units is often a bigger problem than identifying the enterprise. Many times the local unit is not directly registered in administrative registers but needs to be identified indirectly through different sources and even then the local unit itself is not the relevant statistical unit. That is the local kind of activity unit. Often a specific survey on local units is necessary.

Some administrative sources may be useful. For example employment tax systems may hold information on the location of the employees, which can be used to construct a local unit. Other sources may lists business sites such as databases of business sites for business taxes. These need to be linked to other data sources, so the usual matching processes will be required.

In some countries there are no administrative data available to identify local units, and instead these are established through a special register survey. This can be an expensive option, since the survey will be required to collect a number of variables for each site. These are also very burdensome for businesses. A further option is to amalgamate the register local unit survey, with an annual employment survey, which will minimise the data collection costs for both parties.

Many small businesses will never have more than one site, so it is possible to stratify a register proving survey to ensure all larger businesses are targeted, and smaller businesses are sampled very lightly to clarify the probability of smaller businesses owning more than one site. A modelled site based on the enterprise could then be possible for small businesses.

10.3 Converting to statistical units

The following paragraphs consider the processes that are often necessary to transform data from administrative sources into the characteristics required for statistical business registers. The main processes are listed and explained in brief, although they will need to be adapted to fit the specific requirements of each country and source. A principle adopted by several national statistical institutes is that the statistical business register should serve as the gateway for administrative microdata. This means that all administrative data at the level of individual

businesses or sites should be fed through, but not necessarily stored in the business register. This enables the units and variables in the administrative data source to be matched to those in statistical surveys based on the register. Another possibility is to use the administrative units as observation units and through the links in SBR to transform the administrative data into statistical data at the level of the statistical unit for each individual Enterprise or Local Kind of Activity Unit.

10.3.1 The first steps generally concern processing the administrative data to check its quality and coverage. Preliminary analyses may be useful, or even necessary, to check several points including that:

- The file is the expected size, i.e. it contains the expected number of records and variables.
- The main variables are in valid formats e.g. dates are within a permissible range, text fields contain only text characters, numeric fields contain only numbers, etc.
- The coverage of the main variables, e.g. that identity numbers, addresses, economic activity codes etc. are present for all units.
- The variables are plausible, e.g. numeric variables are within an acceptable range and any codes used are valid.

If the preliminary analyses show that the source is of sufficient quality and may help to increase the quality of the business register, a second group of procedures is required to translate the data to meet the requirements of the statistical business register. These procedures refer to the pre-treatment of the administrative data, including the application of statistical definitions of units and variables.

10.3.2 On line deliveries

In the ideal situation the SBR is updated with information in an on line like situation. In this case it is not possible to validate a file as such, but still the variables have to be validated and the converting to updating the statistical units is still relevant.

In this situation it is necessary to have a day to day follow up on units which do not pass for updating of one or another reason. Some 'observation lists' have to be made.

10.4 Concrete transformation rules

An administrative source is unlikely to use the same definition of units required by a statistical business register, especially in the case of complex businesses. As already mentioned, it is important to find out if it is necessary to have an intermediate step to create a real LeU with the links to different administrative units, eventually received from different administrative sources. It might be difficult to automate the transformation from LeUs to Enterprises in case there is not a one to one relationship. The translation of the administrative legal units into statistical units may therefore involve steps such as the creation of algorithms or look-up tables to convert variables and classifications, e.g. tables that convert economic activity codes to NACE codes and location details into geographic location codes.

The rules are very depending on the concrete situation in each country.

11. Quality improvements and addition of specific stratification variables

11.1 Administrative basic sources

Administrative sources are useful sources for updating existing register variables, e.g. name, address, contact details, size and classification variables. Where the administrative source and the statistical business register do not agree, it should be possible to investigate the reasons,

and thus gain a greater understanding of register quality. Surveys may be used to investigate such discrepancies, either specifically for this purpose or as part of some other data collection exercise. These investigations contribute to deduce which rules might be used when updating data is received from different sources.

Administrative sources often hold variables not normally found in statistical business registers. These may be of use for stratification purposes for certain types of surveys.

11.2 Associated registers

Variables specific to a particular sector of economic activity may be made available by using associated registers, e.g. the variable ‘sales space’ for businesses in the retail trade; the variables ‘category/number of stars’ or ‘number of beds’ for hotels; or import/export figures etc. These registers might be used in data confrontation and will be useful for validating the statistical units.

Associated registers can add value to statistical business registers by adding a wider range of variables both for stratification and analysis purposes.

One factor that greatly facilitates the use of associated registers concerns the rights and mechanics of access to administrative sources for statistical purposes. Many potential associated registers are likely to exist, but they are of no real use if the legal and technical means to link them to the statistical business register are not in place. The statistical office could invest resources to enable links to be made to the statistical business register.

The use of associated registers can increase sampling efficiency (i.e. a reduction in the number of units sampled, whilst still retaining the same level of accuracy of survey results) in two ways:

- The statistical business register includes links to units in associated registers which contain additional stratification variables.
- The statistical business register benefits from the improved quality of variables due to the ability to cross-check this register with external sources.

The use of associated registers can reduce the response burden on businesses either through increased sampling efficiency or by removing the need to conduct a survey altogether. They may also include additional units not present in the statistical business register, thus extending the coverage of the population of interest.

To ensure that associated registers are sufficiently coherent with statistical business registers, it may be useful to consider additional criteria, e.g. common unit identifiers, common definitions and classifications. The greater the coherence, the more useful an associated register is likely to be. In addition, the authentic source and the authentic store for those characteristics which are common in both registers should be determined

12. Statistical sources

Different types of statistical sources include:

12.1 Economic census: Economic or establishment censuses can normally provide a comprehensive set of data for establishing the frame of the universe for economic statistics. Economic Censuses are a well established method of initially building Statistical Business Registers particularly in developing countries, but they do have a number of drawbacks. Economic census data is a good start to establish the Statistical Business Register. However, to conduct a census is a resource-intensive exercise and requires large inputs of manpower and time. Censuses therefore tend to be carried out infrequently, for example, once every five

years. The census is undoubtedly the most useful instrument available when a country is initiating an economic statistics programme. Trained field enumerators can seek out each physically recognizable place of business and collect the necessary information by direct interview and observation. In addition to its high cost, this approach has the disadvantage of not being able to document non-recognizable places of business or enterprises without a fixed location (as example, web-based business or individual persons). Because of its infrequent execution a census has its weakness to update regularly the Statistical Business Register. Hence, once the universe of enterprises and establishments is identified additional data sources (e.g. administrative source, enterprise surveys) have to be associated which are able to maintain this business population (e.g. manage the change of characteristics) to add birth of units and to identify death and inactive units This subject is particularly relevant in the developing countries, where companies' births and deaths get high levels.

12.2 Feedback from enterprise surveys

Feedback from enterprise surveys is a vital tool for creating and updating the business register, as it provides new information on changes in contact address, closure of the business, change in the economic stratification characteristics of the unit, etc. Feedback from enterprise surveys has the advantage that it is already available on statistical unit level (i.e. establishment or enterprise). It is recommended that close contact between statistical compilers and Statistical Business Register staff is established to update the register with survey results. Hence, statistical compilers should play an active role in the data quality management of the register. However, it should be mentioned that registers updated exclusively through enterprise survey results have some essential drawbacks as it will be difficult to maintain the entire business population in the register if the feedback is coming only from samples which are already drawn from the Statistical Business Register.

Also there is the issue of feedback bias, where the feedback is provided from sample surveys. In these cases the register is only updated for selected businesses, which can lead to bias. For example if the retail sales survey feeds back information that certain businesses should not be classified to retail, then these will be moved out of the retail population and into another sub population on the register, but the retail survey cannot identify those that should be moved into the retail population. CHECK ????

12.3 Statistical Business Registers surveys

Register updating information that cannot be obtained from survey feedback or from administrative source on which the register is based could to be obtained by register update surveys (sometimes termed nature-of-business surveys) and profiling operations conducted by business-register staff. Control surveys specially designed to measure register accuracy allow the measurement of errors in classification by activity or size and an estimation of the proportion of falsely active units. It is important, however, to deal fully with the question of non-response in order to find out whether units that do not respond have ceased trading or whether they are omitting or refusing to reply to the survey.

One possible way to keep the register more up to date with no administrative source would be to make samples every year where the biggest units are covered 100%, medium sized units are covered e.g. 50% and small units e.g. 10%. The problem isto find the new units.

In case no administrative sources are available, it might be necessary to pick out smaller or bigger areas to find out the coverage and the quality oft he SBR. Of course it will take time to cover all areas.

12.4 Business Profiling

Business profiling is the practice of using company accounts to build the structure of mostly large complex groups. These structures are used to produce a reporting structure for the surveys conducted by the NSI. Profiling usually involves contact with the business being profiled, to build a good understanding of the structure, however it is possible to employ some of the techniques, by just using the published accounts.

While profiling is not a primary source, it does provide valuable information on the larger and more complex businesses that are often vital in their contribution to the GDP of the country.

More details of the practical profiling process will be included in chapter **XX**.

13. Combining Administrative and Statistical sources

In practice to build a comprehensive business register, a combination of administrative and statistical sources can be used. For example administrative sources can be used to identify legal units and convert these to enterprises, while in most cases local units/LKAUs could be identified by a survey of the enterprises.

The idea is to find areas where information from one source does not match information from another source.

One example could be that only one LKAU has been identified to an Enterprise, but at the same time information on employees indicates that for instance half of these live in an area which is far from the identified LKAU. This indicates that either one LKAU is missing or that the employees are linked to the wrong Enterprise.

Another example is that information received on turnover from VAT and connected to an Enterprise does not correspond to the calculated turnover based on information received on import/export. A third example could be receiving information in a survey that the activity is no longer carried out by LeU which has been asked in the survey, but a new LeU has taken over the activity. A taken over has been identified.

14. Links to other types of registers

There are benefits in linking the statistical business register to other types of business register. For example overseas trade or agricultural administrative registers. It is possible to improve the coverage or hold extra variables by linking to other registers. So these can play a part in the building of the Statistical Business Register. These links may also provide opportunities for data analysis by linking for example trade records by employment sizeband.

15. Other and New types of data sources

As well as the traditional sources listed above, new opportunities are emerging. These are associated with increased computer power and on-line filling to administrative systems. These include:

Payroll systems

Where Payroll service providers undertake a payroll management service to business. These involve paying client staff and making returns to the revenue department on their behalf. It may be possible to build agreements with the provider and client to provide data to the statistical office. This saves the business from the burden on survey questionnaires and provides a fast and efficient flow to the statistics office. The identification of the businesses is required to link to the business register, to pick up additional information such as classification.

Accounting system data

Similarly accounting software systems can be exploited to provide data to statistical offices, with the agreement of business users. Again the main incentive for the business is to reduce the burden of completing survey questionnaires.

Internet search

This may be useful as a supplementary manual tool, but cannot currently be used for identifying the business population ???

Commercial data providers

There are a number of commercial data providers businesses that provide global, regional and domestic company information. They serve mainly businesses in credit decisions, marketing etc. However, those data may be also valuable for maintaining business registers in particular on enterprise group structures. It should be noted that those data are by definition based on publicly accessible information and could be therefore obtained also directly by the statistical authority. However, collection matching and processing within the national statistical institute might be too burdensome and acquiring tailor-made data from commercial providers could make the maintenance of the Statistical business register

Other

Work of Mexico using Google Earth could be an example of what can be done ???
The need for meeting the information demands of the society in general have been the basis for new shares of some NSOs, as the case of the National Institute of Statistics and Geography (INEGI) of Mexico which, as mentioned in chapter 2, has made the information of its statistical business register, DENUE, available to all types of users through a consultation system that gathers the geographical information INEGI produces, through a platform with a geographic information environment, allowing the users to perform analysis on the economic activity related to its geographical distribution.

This information conjunction and the use of a geographic information system enable users for example, to perform analysis on economic activities by geographic area, facilitating the identification of clusters, support the planning, designing and assessing the programs to foster productive activity by specific areas, measure the impact of natural disasters in economic areas in the country, detect the needs of infrastructure development, etcetera.

DENUE's consultation system of Mexico, allows the user to select the features of the economic units needed or wished to know: the economic activity, the size (based on employed personnel), and the geographical area. Once the sub-universe of interest has been selected, the system provides the economic unit's information that meets the criteria defined and displays the location of the selection on cartography developed by the National Institute of Statistics and Geography. DENUE's consultation system provides the user with a friendly tool that enriches the information with historical images and street level view.

Mexico's experience with DENUE's consultation system, that was published on INEGI's web-site (<http://www3.inegi.org.mx/sistemas/mapa/denue/default.aspx>) has been of great value, because it indiscriminately and simultaneously provides all kinds of domestic and foreign users with data on the economic units in the country, meeting demands of specialized information as well as those with no statistical purposes.

Annexes

A table showing the core variables in the SBR and its most likely administrative sources.

Annex X: An example of How to use administrative registers and a possible development

- A short presentation of the situation (listing of the development)
- The first establishment of the Danish Business Register
- A presentation of the three/four main data sources (VAT, Tax-register, wage earners' supplementary pension, (register of companies was not used) – and address register – what are the main purposes of the registers, what)
- The work to establish a Business Register (common for sale and for statistical purposes with a clear distinction between the public part and the statistical part; form an ID that could be used as identifying the LeU)
- The work to establish LKAUs
- The work to establish a coherent Business Register
- The work to get information on turnover and employees
- The second establishment of the Danish Business Register
- The corporation with the administrative sources – the authorities recognition of the advantages to have a fully common and public Business Register with the responsibility of the LeUs to update information
- Mm
- Updating frequency
- Quality management
- Data confrontation

References

Business Registers Recommendations Manual. Eurostat 2010

Draft Working Paper on Statistical Business Registers. UNSD and Eurostat June 2011