

**BUSINESS REGISTER  
RECOMMENDATIONS MANUAL**

**Chapter 20  
The Use of Administrative Sources**

# THE USE OF ADMINISTRATIVE SOURCES

## Contents

- 20A – Introduction
- 20B – Definitions of Administrative Sources
- 20C – Why Use Administrative Sources?
- 20D – Problems Using Administrative Sources
- 20E – Quality and Administrative Sources
- 20F – Using Administrative Sources in Practice
- 20G – Access to Administrative Sources
- 20H – Satellite Registers
- 20I – The Example of Intrastat Registers
- 20J – Relation to the BR Regulation

## 20A - Introduction

20.1 This chapter describes the relationship between administrative data sources and statistical business registers, focusing on how those sources can be used to maintain and enhance the register. The definition of administrative sources is considered, along with the expected benefits and problems relating to their use. The specific case of registers that can be linked to a statistical business register (known as "satellite registers") is also considered, focussing on the example of Intrastat data.

## 20B - Definitions of Administrative Sources

20.2 A traditional definition of administrative sources is that they are files of data collected by government bodies for the purposes of administering taxes or benefits, or monitoring populations. This narrow definition is gradually becoming less relevant as functions previously carried out by the government sector are, in many countries, being transferred partly or wholly to the private sector, and the availability of good quality private sector data sources is increasing. It is therefore necessary to consider a wider definition of administrative sources.

20.3 For the purposes of this manual, administrative sources are therefore defined in the widest possible sense:

*Administrative sources are sources containing information that is not primarily collected for statistical purposes*

This wide definition allows the consideration of virtually the whole range of non-survey inputs to business registers, regardless of origin.

20.4 Eurostat collects information on the administrative data sources used for statistical business registers in the annual business registers questionnaire. 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. Information is also obtained in some cases from published accounts, private sector business data providers, or utility company records. In some cases, different sources are used for specific categories of units defined by criteria such as size, economic activity or legal form. These sources can in certain circumstances be used to create "satellite registers". These are discussed in more detail in section 20H.

20.5 Some countries are exploring the possibility of automatically extracting data from the internal financial or management accounting systems of businesses. Under the wider definition set out above, this should also be considered to be use of data from an administrative source.

## **20C - Why Use Administrative Sources?**

20.6 The following paragraphs explore the reasons for using administrative sources, focussing on the advantages of this type of data. The use of administrative sources also raises a number of problems, which are considered, including possible solutions in section 20D.

20.7 The main advantage of administrative sources over survey data is usually cost. Surveys are very expensive, particularly if they are conducted as censuses, or involve the use of personal interviews. Administrative sources are often "free", particularly if they originate from the public sector. If there is a charge for the administrative data, it is often still cheaper than collecting the same information via a survey. Less 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 unlikely that they can be satisfactorily populated and maintained solely by survey data.

20.8 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 institution. A related 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.

20.9 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 point of view of statistical business registers, given that such registers should cover at least all businesses with a labour input of one person half time or more (see chapter 6).

20.10 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.

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

20.11 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. See paragraph 20.18 below.

20.12 Public opinion relating to the sharing of data, particularly between different government departments, varies considerably from country to country. Where public opinion generally accepts, or is in favour of data sharing, the increased use of existing data sources can help to enhance the prestige of a national statistical institution by making it seem more efficient and cost-effective.

## **20D - Problems Using Administrative Sources**

20.13 Although there are many good reasons for using administrative sources, there are also a number of common problems. 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. One specific problem, that of getting access to administrative sources in the first place, is treated separately in section 20G below.

20.14 Paragraph 20.12 considered how public opinion might favour the sharing of data in some countries. In other countries, however, there may be public unease at the thought of data being shared around 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). The publication of analyses of the costs and benefits, both to government and to businesses, of the use of different sources may also help.

20.15 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. The process of converting from administrative units (which may often be equivalent to legal units) to statistical units (i.e. enterprises and local

units) is known as profiling, and is discussed in chapter 19, along with rules for combining legal units into enterprises.

20.16 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 have 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.

20.17 Similarly, the classification systems used within administrative sources may be different to those used in the statistical world. 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 correlation tables 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.

20.18 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 they 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 are 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.

20.19 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 get early warning of possible changes, and by drawing up contingency plans.

20.20 If data from several administrative sources are 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, and the need for clerical investigation of possible matches.

20.21 Another problem where multiple sources are used concerns consistency between the sources. Data from one source may appear to contradict those 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 are 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.

## **20E - Quality and Administrative Sources**

20.22 The issue of quality in relation to statistical business registers is discussed in detail in chapter 10, but it is worth commenting on the assessment of the quality of various actual and potential administrative sources used to maintain statistical business registers. If the manager of such a register is fortunate enough to be faced with a choice of two or more administrative sources, how can he or she determine which source has the higher quality?

20.23 There is no simple answer to this question, but there are several possible approaches, which may be used in combination. It can be useful to compare sources in terms of coverage and accuracy of the variables. This can be aided by some sort of quality survey to determine the correct values of certain variables.

20.24 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 are collected and processed. Thorough understanding of a source will allow a more accurate assessment of strengths and weaknesses.

## **20F - Using Administrative Sources in Practice**

20.25 The following paragraphs consider the processes that are often necessary to transform data from administrative sources into the variables required for statistical business registers. The main processes are listed and briefly explained below.

- ***First steps and the transition from administrative to statistical data***

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

- The main characteristics of the administrative data e.g. date of reference, the total number of units, procedures used to update the source, etc.

- The coverage of the main variables e.g. identity number, address, economic activity codes, number of employees, sales space, products sold etc.
- The error profile of these variables, especially the variables needed to determine continuity of statistical units (e.g. identity numbers, economic activity codes, location, dates, etc.)

20.27 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 are 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.

20.28 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. The translation of the administrative data into statistical data 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 NUTS codes.

- ***Linking and matching***

20.29 The next step is to identify the link between units in the administrative source and the corresponding units in the statistical business register. If available in the administrative source, the basic link can be established through an administrative code (e.g. fiscal code or VAT number) already recorded in the business register.

20.30 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, and is intended to be discussed in more detail in a later chapter. With this approach it is, however, possible to link units in error (sometimes known as false matches).

20.31 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.

20.32 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.

- ***Quality improvements and addition of specific stratification variables***

20.33 Administrative sources can provide a useful check for 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 can be used to investigate such discrepancies, either specifically for this purpose, or as part of some other data collection exercise.

20.34 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. Such sources may be used to develop "satellite registers", which are discussed further in section 20H.

## **20G - Access to Administrative Sources**

20.35 Issues relating to gaining access to administrative sources are dealt with thoroughly in the Eurostat publication "Use of Administrative Sources for Business Statistics Purposes - Handbook of Good Practices", so are only summarised and related specifically to statistical business registers in this chapter.

20.36 There are two aspects to getting 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. The practical issues concerning access to administrative sources also vary from country to country, but here it is more appropriate to consider good practices and how they can be spread.

20.37 The annual business register survey conducted by Eurostat has shown that the main administrative sources used for statistical business registers are tax records (e.g. those relating to VAT or employee income tax), or details of compulsory registration with bodies such as chambers of commerce. Access to such public sector sources is usually free, but may be difficult to obtain due to policy, legal or resource considerations. Private sector administrative sources are used in some countries, e.g. for information on links and ownership. Access to these sources is usually easier to arrange, but is normally subject to a charge.

20.38 It is recommended to have some sort of formal agreement 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.

20.39 Alongside 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 investing some time to visit 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 identification and resolution of problems as they arise, and before they escalate.

## 20H - Satellite Registers

20.40 One way of using administrative data in practice is to organise these data into specific registers linked to the statistical business register. If these registers meet certain criteria, they can be referred to as "satellite registers". Satellite registers are therefore defined as registers that are available to the national statistical system, contain information about businesses, and fulfil the following conditions:

- They are not an integral part of the statistical business register as referred to in the business registers Regulation, but are capable of being linked to it.
- They are more limited in scope than the statistical business register, e.g. in terms of NACE, but within that scope they may have more extensive coverage of units and/or variables.
- They contain one or more variables that are not found in the statistical business register. Such variables are generally capable of being used for stratification purposes.

20.41 Thus satellite registers can be constructed using information from administrative sources, statistical surveys, or a combination of both. In some cases they may add, combine or otherwise transform variables, though in others they may be more or less identical to a particular source. To ensure that satellite 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 a satellite register is likely to be.

20.42 The scope of satellite registers can be determined by:

- Economic activity - they may contain businesses with similar activities e.g. retail trade, accommodation, road haulage etc..
- Size - They may contain units with a certain number of employees or turnover over a certain level, e.g. they may contain the subset of "large enterprises"
- Characteristics - They may contain units with a common characteristic, e.g. those that engage in foreign trade (units on the register of intra-Community operators - variable 1k in the business registers Regulation)

20.43 Variables specific to a particular sector of economic activity may be made available by using satellite registers, e.g. the variable "sales space" for businesses in the retail trade; or the variables "category/number of stars" or "number of beds" for hotels. Satellite registers can add value to statistical business registers by adding a wider range of variables both for stratification and analysis purposes.

20.44 One factor that greatly facilitates the use of satellite registers concerns the rights and mechanics of access to administrative sources for statistical purposes. Many potential satellite 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.

20.45 The use of satellite 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 additional stratification variables via satellite registers.
- The statistical business register benefits from the improved quality of variables due to the ability to cross-check this register with external sources.

20.46 The use of satellite 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.

## **20I - The Example of Intrastat Registers**

20.47 The following paragraphs consider the potential links between statistical business registers and registers of intra-Community operators (Intrastat registers), and the possible benefits deriving from these links. An Intrastat register can be considered as a form of satellite register. This example is of particular interest because the business registers Regulation includes a requirement for a specific variable to be held at legal unit level: “Reference to the register of intra-Community operators drawn up in accordance with Council Regulation (EEC) No 3330/91 of 7 November 1991 on the statistics relating to the trading of goods between Member States” (variable 1k).

20.48 Council Regulation 3330/91 defines the rules for the compilation of statistics relating to the trading of goods between Member States. According to this Regulation, Member States should have had a register of Intra-community operators at their disposal by 1 January 1993. This Regulation was drawn up at the end of 1991, two years before the registers Regulation, therefore the registers Regulation requires a link to the Intrastat register at legal unit level.

20.49 Articles 5 and 6 of Commission Regulation 1901/2000, which implements and amends Council Regulation 3330/91, define a list of minimum data fields to be recorded in the Intrastat register:

- 
- Full name of the person or firm
- Full address including postcode
- VAT registration number
- The year and month of entry in the register
- Where applicable, whether the operator is a party responsible for providing information, or a declaring third party, upon either consignment or receipt
- In the case of a party responsible for providing information; the total value of his intra-Community operations by month and by flow, together with the value of trading of goods between Member States which the operator has mentioned in his periodic tax declaration.

20.50 The links between the statistical business register and the Intrastat register can be analysed according to the following elements:

- Sources of information - Fiscal registers are the main source of information of the Intrastat register, and are usually an important source for statistical business registers.
- Reference units - The registers Regulation identifies three reference units for statistical business registers, the legal unit, the enterprise, and the local unit. The Intrastat Regulation does not define any unit of reference, however, as fiscal registers are the main source of information for Intrastat registers it can be assumed that the unit of reference is generally the legal unit, that is every legal unit responsible for the control of the operator having Intra-Community import/export movements.
- Variables - Commission Regulation 1901/2000 specifies the variables to be included in the Intrastat register. Among these, name and address of the person or firm corresponds to variable 1b of the legal unit in the statistical business register. Another possible common variable is the VAT registration number, which can also appear as variable 1j in the statistical business register.

20.51 Given these similarities, there should be a strong link between the Intrastat register and the statistical business register. Conceptually, the Intrastat register represents a logical subset of the business register corresponding to the legal units with intra-Community exchanges. The subset should be identified via variable 1k of the legal unit in the statistical business register, which should be present by definition in the Intrastat register. If this variable is missing, it may still be possible to establish a link using another identification number e.g. the VAT registration number, or by matching based on name and address.

20.52 It should be noted that there may be some units in the Intrastat register which are not included in the statistical business register, e.g. units classified to NACE sections A, B and L, which are not currently compulsory for statistical business registers, though are included on an optional basis by many Member States.

20.53 Both the Intrastat register and the statistical business register can benefit from the link between them. Specifically, the statistical business register has access to a reliable tool for checking the quality of its own data regarding a limited (in number), but important subset of enterprises. The link between the Intrastat register and the statistical business register may also help to identify inconsistencies, e.g. if the product code of the exchanged goods declared by intra-Community operators is available, this can be compared with the activity code of the statistical business register.

## **20J - Relation to the BR Regulation**

20.54 Article 7 of the business register Regulation states that "Each national statistical institute shall be authorised to collect for statistical purposes information covered by this Regulation which is contained in the administrative or legal files

compiled on its national territory, in accordance with the conditions determined by national law". The use of administrative sources, as described in this chapter is therefore considered to be an interpretation of the Regulation. The methodological recommendations concerning the transfer of data from administrative sources, and the creation of satellite registers, go beyond the Regulation.

20.55 Legal unit variable 1k, "reference to the register of intra-Community operators drawn up in accordance with Council Regulation (EEC) No. 3330/91 of 7 November on the statistical relating to the trading of goods between Member States" is compulsory in the business register Regulation. Therefore a link between the business register and Intrastat register is required. The text above concerning this link can be considered as guidelines to aid the implementation of the Regulation rather than as a strict interpretation of the Regulation.