United Nations

Department of Economic and Social Affairs

Guidelines on Statistical Business Registers

Department of Economic and Social Affairs
Statistics Division

Studies in Methods   Series F No. 122

Guidelines on Statistical Business Registers

United Nations
New York, 2024

Department of Economic and Social Affairs

The Department of Economic and Social Affairs of the United Nations is a vital interface between global policies in the economic, social and environmental spheres and national action. The Department works in three main interlinked areas: (i) it compiles, generates and analyses a wide range of economic, social and environmental data and information on which United Nations Member States draw to review common problems and to take stock of policy options; (ii) it facilitates the negotiations of Member States in many intergovernmental bodies on joint courses of action to address ongoing or emerging global challenges; and (iii) it advises interested Governments on the ways and means of translating policy frameworks developed in United Nations conferences and summits into programmes at the country level and, through technical assistance, helps build national capacities.

Note

The designations employed and the presentation of the material in the present publication do not imply the expression of any opinion whatsoever on the part of the United Nations concerning the legal status of any country or of its authorities or the delimitations of its frontiers. The term "country" as used in this publication also refers, as appropriate, to territories or areas. The designations of country groups in the publication are intended solely for statistical or analytical convenience and do not necessarily express a judgment about the stage reached by a particular country, territory or area in the development process. Mention of the names of firms and commercial products does not imply the endorsement of the United Nations. The symbols of United Nations documents are composed of capital letters and numbers. Mention of such a symbol indicates a reference to a United Nations document.

Preface

At its forty-eighth session, in March 2017,[1] the Statistical Commission reaffirmed the importance of statistical business registers, underlined its important role in the production of economic statistics, and stressed the continued need for capacity-building activities on statistical business registers in developing countries. The Commission endorsed the proposal to draft and publish *Guidelines on Statistical Business Registers* based on existing regional guidelines, notably the Guidelines for Statistical Business Registers of the Economic Commission for Europe, and to broaden the scope of the guidelines by addressing the specific needs of countries with less developed statistical systems. In particular, the Commission requested that the global guidelines incorporate practical guidance and country cases applicable to a broad range of statistical systems.[2] Furthermore, the Commission entrusted the preparation of the global guidelines to the body subsequently established as the Committee of Experts on Business and Trade Statistics.

The present Guidelines are the result of a preparation process which started with the identification of a list of issues that needed to be updated from the Guidelines for Statistical Business Registers of the Economic Commission for Europe. The list included issues relevant for developing and maintaining statistical business registers in less developed statistical systems, characterized by a large informal sector and data-collection practices using business surveys to complement the limited availability of administrative data. The issues also included the choice of statistical units and the role of multinational enterprises in the context of globalization. This list was prepared by the Committee of Experts through an online consultation and further discussion during the Committee's first meeting in Mexico in June 2018 and included in the summary report of that meeting.[3] Contributions from experts were collected and discussed during an ad hoc meeting organized back-to-back with the twenty-sixth meeting of the Wiesbaden Group in September 2018.

The draft United Nations Guidelines on Statistical Business Registers were circulated for global consultation during the period from January to February 2019. Comments received during the consultation were carefully reviewed and included in the final version. The Guidelines were endorsed by the Statistical Commission at its fiftieth session.[4] The Commission also encouraged countries to use the Guidelines when establishing and maintaining their statistical business registers.

Statistical business registers are the backbone of the statistical system. They play a central role in the production of official statistics, both in terms of the way in which the statistics are produced and in terms of the quality of the statistics that draw on information from them. The Guidelines reflect the global recognition by the international statistical community of the importance of statistical business registers and they will provide the basis for the formulation of technical assistance programmes to strengthen statistical business registers in countries.

[1] See report in document E/2017/24-E/CN.3/2017/35.

[2] Ibid., decision 48/114.

[3] See summary report at https://unstats.un.org/unsd/trade/events/2018/mexico/Summary%20Report%20of%20the%20UN%20Committee_Mexico%20meeting_final.pdf.

[4] See report in document E/2017/24-E/CN.3/2017/35, decision 50/111.

Acknowledgements

Abbreviations and acronyms

ABR	administrative business register
ABS	Australian Bureau of Statistics
ADB	Asian Development Bank
AHV	Swiss social security number
ASIA	Business Register of Active Enterprises (Italy)
BDR	Building and Dwelling Register (Denmark)
BKPM	Indonesian Coordinating Investment Board
BPM6	Balance of Payments and International Investment Position Manual, sixth edition
BPS	Badan Pusat Statistik (Central Bureau of Statistics in Indonesia, also known as Statistics Indonesia)
BRN	busines registration number (Mauritius)
BURNR	local unit identification number (Switzerland)
CAPMAS	Central Agency for Public Mobilization and Statistics (Egypt)
CBRD	Corporate and Business Registration Registration Department (Mauritius)
CBS	Central Bureau of Statistics, also known as Statistics Netherlands
CEMPRE	Central Business Register (Brazil)
CNSS	National Social Security Fund (Tunisia)
CPC	Central Product Classification
CPR	Central Person Register (Denmark)
CSBR	Canadian Statistical Business register
DENUE	National Directory of Economic Units (Mexico)
DGI	General Directorate of Taxes (Tunisia)
ECB	European Central Bank
ECE	Economic Commission for Europe
EFTA	European Free Trade Association
EG	enterprise group
EGR	EuroGroups Register
EIN	establishment identification number
ENT	national enterprise
ESA	European System of Accounts
ESBRs	European System of Interoperable Statistical Business Registers
ESS	European Statistical System
FATS	foreign affiliates statistics

FDI	foreign direct investment
FSO	Federal Statistical Office (Switzerland)
FTEs	full-time equivalents
GDP	gross domestic product
GEG	global enterprise group
GEN	global enterprises
GEN/ENT	global enterprise versus national enterprise
GEO	geographical unit
GEOSTAT	National Statistics Office of Georgia
GIS	geographical information systems
GLEIF	Global Legal Entity Identifier Foundation
GSBPM	Generic Statistical Business Process Model
GSIM	Generic Statistical Information Model
GVC	global value chain
IBGE	Brazilian Institute of Geography and Statistics
IBR	Integrated Business Register
IDSR	Inter-Departmental Business Register (United Kingdom)
IES	Simplified Business Information (Portugal)
IFRS	International Financial Reporting Standards
ILO	International Labour Organization
INEC	National Institute of Statistics and Census (Costa Rica)
IMF	International Monetary Fund
INEGI	National Institute of Statistics and Geography (Mexico)
INS	National Institute of Statistics (Tunisia)
INSEE	National Institute of Statistics and Economic Studies (France)
ISIC	International Standard Industrial Classification of All Economic Activities
ISO	International Organization for Standardization
ISSF	integrated statistical systems framework
Istat	National Institute of Statistics of Italy
IT	information technology
KAU	kind-of-activity unit
LBU	large business unit
LEI	legal entity identifier
LeU	legal unit
LKAU	local kind-of-activity unit
LSS	large sample survey
LU	legal unit
MNE	multinational enterprise group
MRA	Mauritius Revenue Authority
NACE	Statistical Classification of Economic Activities in the European Community

NAPR	National Agency of Public Registry (Georgia)
NOGA	General Classification of Economic Activities (Switzerland)
NPI	non-profit institution
NPISH	non-profit institution serving household
OECD	Organisation for Economic Co-operation and Development
OPU	operating unit
OU	observational unit
PAYE	pay-as-you-earn
PCBS	Palestinian Central Bureau of Statistics
QIU	quality improvement unit
RAIS	Annual Report of Social Information (Brazil)
RDBMS	relational database management system
REE	Register of Enterprises and Establishments (Costa Rica)
RENEM	Statistical Business Register of Mexico
RMSE	root-mean-square error
RNT	National Tourism Register (Mexico)
SBER	Swiss Business and Enterprise Register
SBR	statistical business register
SBS	structural business statistics
SDMX	Statistical Data and Metadata eXchange
SEEA	System of Environmental-Economic Accounting
SIC	standard industrial classification
SID	Directory Information System (Colombia)
Sirus	Système d'immatriculation au répertoire des unités statistiques (French information system of statistical units)
SMA	subject matter area
SME	small and medium-sized enterprise
SNA	System of National Accounts
SPE	special purpose entity
SSR	social security register
SQL	Structured Query Language
TEG	truncated enterprise group
TEN	truneconocated enterprises
VAT	value added tax

Contents

Boxes

Figures

Chapter 1
Introduction

1.1 Context

1.1.   In considering national and international economic policy actions, statistical data are an important input for analysts, researchers and politicians. In addition, the general public are interested in the status and development of the economy as they form part of the economic system as consumers, entrepreneurs, employers or employees. For that reason, the production of high-quality official economic statistics is an important task of national statistical offices. Business and economic statistics cover a broad area of statistics, ranging from economic censuses and other structural information to production statistics, foreign trade and foreign affiliates statistics (FATS), research and development and innovation statistics, and also business cycle indicators, production indices, price indices and various other types of statistical domains. Added to these, the production of national accounts data for analysis from a macroeconomic point of view is part of the economic statistical system of a country.

1.2.   The demand for economic statistics is growing in all countries and also at the international level. To a large extent this is the result of the increasingly complex nature of economies, with a growing division of labour at both the national and global levels. To remain relevant and to meet user needs, economic statistics should, therefore, cover new areas and provide more detailed information about activities and variables. There is also a general need to improve timeliness, coherence and reliability at both the national and global levels.

1.3.   While business statistics should achieve these new and enlarged goals, it should do so in an efficient way and with a response burden that is as light as possible. This means that efforts must be made to modernize the production processes, to use administrative data to reduce survey burden, to redesign survey systems, to harmonize surveys and variables, comprehensively to link administrative data and survey sources, and to cooperate more closely with administrative authorities on a continuous basis to improve the quality of the statistical business register (SBR). Better consistency between the various statistical areas at the national and international levels has also become a much more important goal than it was in the past, when economic development was less affected by international dependencies and production and markets were much less globalized.

1.4.   In all these developments of business statistics the SBR plays a key role. Accordingly, the challenges of current and future business statistics are at the same time also challenges for the SBR. The SBR delivers the basic information for conducting economic surveys by providing the populations of statistical units and their characteristics. In addition, the SBR provides links to administrative units and registers, thus enabling the use of administrative data for statistical purposes. It also provides unique identifiers enabling linkages at the micro level across statistical domains as needed.

1.5.   High-quality business statistics depend on a high-quality SBR. A high-quality SBR fulfils the user needs in an optimal way, and is based on international concepts, definitions and classifications. Thus, it serves as the basis for international harmonization of economic statistics in terms of coverage, statistical units and frame methodology. The main users of SBR information are survey staff in the national statistical offices. These staff use the SBR as the source of input for survey design and survey operations. As SBR and survey staff are members of the same organization, close cooperation and exchange on user needs should be easily achievable. The basic roles that the SBR plays for the production of business statistics has a significant effect on the national accounts, which need high-quality and consistent business statistics to achieve their goals. Accordingly, the SBR plays an important role in the implementation of the concepts and methods of the System of National Accounts (2008 SNA).

1.2   What is an SBR?

1.6.   An SBR is a *regularly updated, structured database of economic units in a territorial area, maintained by a national statistical office, and used for statistical purposes.* The terms used in this definition are further elucidated below.

Structured database

1.7.   An SBR is a complex structured electronic and relational database, with defined characteristics stored for each economic unit. In circumstances where there is more than one kind of economic unit (which is normally the case) the relationships between these units are also included. For example, for each local unit there is a link to the enterprise to which this unit belongs; for each legal unit there is a link to the enterprise with which it is associated. Economic units are assigned unique numerical identifiers so that they are easily identified and to make sure that no units are included twice.

Economic units

1.8.   Economic units comprise legal-administrative units and statistical units. Legal-administrative units are units registered in administrative registers, such as taxation registers, social security registers, company registers, registers of the chamber of commerce, and so forth, including also non-market units such as government departments and non-profit institutions (NPIs). Legal-administrative units serve as the basis for the delineation of statistical units, by which they are represented for statistical purposes. Examples of statistical units are the enterprise, the local unit and the establishment. In most cases, a legal-administrative unit corresponds to a statistical unit. In specific cases, however, a legal unit might not correspond to a statistical unit. Accordingly, in an SBR database these two types of units are recorded as separate but related types of entities. This approach applies to all types of units. For instance, enterprises and their local units are recorded as separate types of entities. Even where an enterprise has only a single local unit (which is the most common situation), two units are to be recorded in the SBR: an enterprise unit and a local unit.

Regularly updated

1.9.   An SBR that was not regularly updated would soon lose its value, as ongoing changes in the numbers and structure of the economic units would not be incorporated into the database. Such changes include: new economic units being created and others being closed; economic units being merged or changing legal form, location or

activity; units having higher or lower turnover over time or engaging more or fewer employees than in the past period. Thus, the stratification variables used for the delineation of survey frames may also change over time.

1.10.   The frequency of SBR updating depends on data availability, including the availability of administrative data. If data for SBR updating are only available on an annual basis, the use of the SBR for monthly or quarterly surveys is hampered. Some administrative data are made available on a monthly or quarterly basis, others only on an annual basis and, while some data may be available shortly after the reference period, others may be available only later. A further issue is the time lag between an actual business event and the date on which it is recorded in the administrative or other database. SBR updating is usually a continuous process in order to minimize time lags in processing data. The SBR contains the latest instance of the data collected (with a time stamp). In cases where the SBR is derived primarily from survey or economic census data, the SBR should be updated as soon as data are collected and processed, even if there is a considerable time lag between the data collection and the availability of the final data.

Territorial area

1.11.   An SBR is established to cover all business units that are resident in a specific territorial area. The territorial area is normally the economic territory of a country. Accordingly, an affiliate of a resident enterprise in a foreign country would not be recorded in the SBR of the country of the enterprise. This affiliate would also not be covered in the administrative sources and therefore information for updating would not be available. Information on foreign affiliates, however, might be collected by a specific statistical survey.

1.12.   The backbone role of the SBR is not easily achieved if a separate SBR is established and maintained for each region of a country. This would lead to various practical and conceptual problems. For example, an enterprise could be active in different regions. In order to avoid such problems, it is highly recommended that there should be only one SBR in a country covering all units resident in that country. If regional SBRs exist they must be coordinated at the national level.

1.13.   Restriction to the national territory has the consequence that multinational enterprise groups (MNEs) are not covered in their entirety: only their national parts are covered. The coverage of multinationals would need a supranational SBR with appropriate cross-country cooperation between the national statistical offices.

Maintained by a national statistical office

1.14.   Administrative registers are maintained by administrative authorities for purposes of public administration and programmes. Their coverage, characteristics, and methods of updating are designed to serve administrative purposes. An administrative business register might not cover all units but only selected ones, for example, those with a particular legal form (such as companies), or those with a particular type of economic activity (such as farms), or those in a certain region. An SBR has to fulfil the requirements for economic statistics. The coverage requirements are often quite different and more comprehensive than for administrative registers. Thus, an SBR has to be established and maintained by a national statistical office, in fulfilling its task of producing official statistics.

Statistical purposes

1.15.   An SBR is established in order to provide the frame population for economic surveys and for other statistical purposes. The requirements for an SBR can therefore be quite different from the requirements of an administrative register. Even if the SBR is based on information from administrative registers, the concepts, characteristics, methods of maintenance, and other aspects of the SBR need to be based on statistical concepts, preferably those which are internationally recommended. SBRs also have one other important feature that distinguishes them from administrative and SBRs. The SBR is only to be used for statistical purposes. It may not be used for other purposes, in particular administrative purposes. The reason for this is the confidentiality of individual data collected by national statistical offices. Data on individual natural persons or businesses collected by national statistical offices for statistical compilation are to be strictly confidential, and used exclusively for statistical purposes.

1.16.   In satisfying the requirements implicit in the above definition, an SBR is a complex system in terms both of structure and of technical implementation. It might seem that a quite simple system, such as a list of businesses with the appropriate variables, could serve to provide sets of units for survey frames for collection of data. This would be the case when only one type of unit, for example enterprise or establishment, was of interest, and there were very few surveys. Even then, however, such a list would have to include many variables, such as name, address, legal form, economic activity and other classification codes, and its handling and updating would be difficult tasks. Where there are several types of units and many surveys to consider, a simple list is impractical; a database is needed.

1.17.   A general principle is that a data update to the SBR should be time-stamped to indicate, first, the date when the change has been made, and, second, the date when the change becomes effective. For instance, a merger of two enterprises might have taken place – and be effective from – 1 January 2020, but the appropriate changes to the SBR might be made at a later date, say 1 June 2020, when the information becomes available for SBR updating purposes.

1.18.   A second principle is that SBR data should never be physically deleted. For example:

> ‣ An enterprise that has ceased activity will not be deleted; rather the "activity status" of the enterprise will be set to "inactive".

> ‣ A change of the location of an enterprise will be recorded in such a way that the old address is marked as "no longer valid" and the new address is introduced with an effective starting date.

1.19.   In addition, for a selection of variables the previous values should be stored, either directly in the SBR (in the same file) or in separate historical files, indicating the name of the variable, its previous value, source, and effective dates, or by other information technology (IT) tools.

1.20.   To manage requirements of these kinds a well-structured relational database is needed. They cannot be managed as a list.

1.21.   As with any other statistical product, quality of the SBR is of extreme importance, in particular since the SBR is not an end in itself, but an input to all the economic statistical products based on it. The quality of an SBR can be measured by the same criteria that are used for measuring the quality of statistical data: relevance, accuracy, punctuality, accessibility, comparability, and coherence. More specifically this means: that the information provided to the users corresponds to their needs; that the economic units covered are in accordance with the defined target population; that

the data are as timely as possible; that the coverage of economic units and their characteristics are as complete and accurate as possible; that the data are comparable over different activities and regions and internationally by virtue of their using the same concepts, methods and definitions; and that the SBR is coherent with other statistical data sets, where relevant.

1.3   Use of an SBR

1.22.   An important purpose of the SBR is to provide sets of units referred to as frames or surveys. A survey frame is a list of units together with the characteristics of these units required in conducting the survey. By providing frames for all relevant surveys from a central source, the SBR is also acting as a coordinating instrument. As the SBR units and their characteristics are updated on a continuous basis, survey frames derived from the SBR can be similarly updated. Updating centrally in the SBR is also much more efficient than leaving it up to the different survey areas to update the survey frames. Thus, a single SBR is a good solution not only for harmonized statistics but also for cost-efficiency.

1.23.   Moreover, the SBR can do more than simply provide frames. It can also be used to combine data from different sources to improve coverage. It can provide a mechanism for identifying units and avoiding double counting; it can facilitate central storage of metadata; and it can coordinate various data streams. All of these help to optimize the quality of statistical processes before direct data collection has even begun.

1.24.   Variations in the coverage, content and quality of SBRs across countries make the comparison of economic statistics more difficult. Accordingly, the coverage and content of SBRs should be harmonized as far as possible and international concepts and definitions should be used to the maximum extent possible. Economic statistics describe economic production processes and financial transactions in various statistical domains. While each domain has its own specific peculiarities, the data collected should be comparable across domains. This is only possible if the populations for the various domains are derived and used in a similar and coordinated way during the course of statistical processing.

1.25.   Using administrative data more extensively for statistical purposes is an important strategic goal for national statistical offices in their efforts to keep up with ever growing demands for economic data by governments, international bodies and researchers, in particular data for microeconomic analysis of small areas, and for specific sectors of activity aggregates. While administrative data may not include precisely the information that the statistician would like to collect and disseminate, they have the virtue of providing more – and possibly almost complete – coverage of a target population, whereas sample surveys cover only a relatively small proportion. The use of administrative data may therefore decrease, or eliminate, sampling errors, significantly reduce or remove non-response, and provide more accurate and detailed estimates for various subpopulations. Thus, to reduce both the costs of data collection and response burden, administrative data sources should be further integrated into the statistical processes used in collecting and compiling economic statistics. The starting point is the use of administrative data in the creation of statistical units in SBR.

1.26.   In some cases, challenges will undoubtedly arise that limit a national statistical office's ability to access useful administrative data. Nonetheless, in implementing an SBR, a profile of the various administrative processes to which businesses are sub-

ject through the various stages of their life cycle should be produced. This includes processes for registration, licensing, hiring (or releasing) employees, paying business taxes (such as payroll taxes, corporate taxes, personal business taxes, value added taxes, import duties, and others) and formally terminating business activities. Having identified these various processes, the feasibility of accessing the administrative data records that correspond to them should then be assessed, most likely by initiating discussions with the various government departments or institutions that administer them. Such consultations might include, for example, tax offices, chambers of commerce, employer registration offices, and other bodies.

1.27.   Owing to the general importance of administrative data for the maintenance of the SBR, it is crucial that national statistical offices have access to such databases. There might, however, be obstacles for such access due to legal issues and also technical and organizational issues. Legal obstacles can be addressed by a review and revision of the statistical laws or other relevant laws. It is important that national statistical offices work with relevant institutions in order to ensure that new and existing administrative data collections are more usable for statistical purposes. Technical and organizational issues hampering access by national statistical offices to administrative databases can be addressed by establishing good cooperation between the national statistical office and the administrative bodies. Such cooperation should be laid down in a memorandum of understanding.

1.28.   New data sources should be evaluated carefully before they are considered ready to supplement or enhance the SBR. They should only be considered as an input to the SBR if they meet the important criteria delineated in section 1.2. In other words, that the source is regularly updated; it meets the SBR definitions of economic units; and it has the quality needed for statistical purposes. Once a data source is implemented it should be regularly monitored for changes and overall quality and a close relationship with the data provider should be developed. This could be covered in the memorandum of understanding.

1.29.   Use of administrative data also increases the range of tasks that can be achieved by an SBR. With these data, the SBR becomes not only a database providing frames for survey purposes but also a powerful database that can be used for the direct production of statistics. Furthermore, links to administrative data and registers facilitate the use of administrative data for supplementing or replacing data collected by surveys. For countries with a large informal sector, SBR improvement surveys (see sect. 1.6 for a definition) and economic censuses will still be necessary to capture information on establishments operating outside the formal sector. In such cases, the use of administrative data could reduce the frequency and scope of SBR improvement surveys and economic censuses.

Box 1.1
Increased use of administrative data: example of Colombia

In business surveys, the use of administrative registers as a complement to statistical surveys has increased. For example, information from the migratory flows provided by Migración Colombia – the immigration authority of Colombia – is used in the statistics of imports and exports of services, and also serves as an input to calculate items related to trips in the country's quarterly survey of foreign trade in services. In addition, the administrative registers of imports supplied by the country's customs agency are used as a source of information for imports of freight services. The information system of the distribution

Box 1.1 (*continued*)

chain of liquid derived petroleum fuels, maintained by the Ministry of Mines and Energy, is used to supplement the fuel information directory of the service stations of the annual trade survey. It integrates the agents of the chain at a national level into a single information system through which the marketing, distribution, transport and storage of liquid fuels derived from petroleum, fuel alcohol and biodiesel are organized, controlled and systematized.

The use of administrative registers can be one of the most useful tools for the generation of statistics. The use that has been made of them does not, however, include certain aspects of quality, management and mechanisms for validation and consistency of information. For example, the Ministry of Commerce, Industry and Tourism is sharing data on the use of hotel registration cards with the National Administrative Department of Statistics in the production of tourism statistics. This practice demonstrates the use of administrative registers for statistical purposes. It also shows how inter-institutional support can ensure the accuracy of collection instruments and shared management of an easy and secure kind conducive to the generation of timely, good-quality and lower-cost information. It will therefore yield more accurate statistics for decision-making.

1.30.   Some harmonization of concepts is needed to make the data usable for the SBR, not only in the case of administrative data, but also when linking and using statistical data sources for the update of the SBR, such as an economic census or household survey. Box 1.2 below illustrates this in the case of household surveys (see also section 3.5.2).

Box 1.2
Issues of linking household and business data in South Africa

For meaningful linkages between household statistics and business statistics, the SBR needs to standardize and harmonize the interpretation of the different types of statistical units. The structural differences between business statistics variables and household statistics variables need to be fully investigated.

A coordinated population register is crucial for the integration of business and household statistics.

In South Africa, the current SBR is a structured database of enterprises registered to pay value added tax (VAT) and indirect tax. This database provides a foundation for creating sampling frames for each economic survey, as it contains the population of statistical units at fixed point in time. Business information contained in the SBR is updated from the South African Revenue Service). The survey of employers and the self-employed and the Quarterly Labour Force Survey cover aspects related to the informal sector. Currently there are no linkages between the SBR and household surveys.

1.31.   New developments in the production of economic statistics will extend the role and use of the SBR. Globalization, the increased use of administrative data, industrialization and emerging new data sources (see box 1.3 for examples of the use of Internet data and big data) are some of the developments that must be taken into account when looking at the future use of the SBR.

[1]   Lotte Oostrom and others, "Measuring the internet economy in the Netherlands: a big data analysis", CBS Discussion Paper, No. 14 (CBS, 2016).

[2]   Steve MacFeely, "Big data and official statistics", in *Big Data Governance and Perspectives in Knowledge Management*, Sheryl Strydom and Moses Strydom, eds. (Hershey, PA, IGI Global, 2019).

[3]   P. Daas and others, "Big data as a source for official statistics", *Journal of Official Statistics*, vol. 31, No. 2 (2015), p. 255.

> **Box 1.3**
> **Examples of the use of data from the Internet and big data**
>
> Statistical offices have been experimenting with the use of big data and Internet-based approaches in various fields of statistics. A broad overview of these initiatives is provided by the work between institutions to develop a new statistical output from the use of new data sources and may be found in the Global Working Group on Big Data for Official Statistics (https://unstats.un.org/bigdata/). A practical example of joint the paper "Measuring the Internet economy in the Netherlands: a big data analysis".[1] Statistics Netherlands (CBS), Google and Dataprovider combine administrative registers with big data to provide a general overview of the use of the Internet economy by Netherlands enterprises. Despite a number of challenges, the use of this kind of information can bring new types of statistics, more granular indicators, reduce response burden and lead to improvements in the timeliness and accuracy of the statistics developed by the national statistical offices.[2] The authors of a study on this question observe that "Big Data can become a useful and reliable data source for the field of official statistics",[3] although there are still some challenges and issues that must be addressed by the national statistical offices.
>
> In addition, examples of the use of new informatics tools to obtain new products include the use of web-scraping, such as in the United Nations project to develop a Global Group Register of the largest global multinational enterprises from information publicly available and the project mounted by the Organisation for Economic Co-operation and Development (OECD) entitled: "Measuring the MNEs using Smart Data: the OECD Analytical Database on Individual Multinationals and their Affiliates".
>
> Another example of the use of web-scraping techniques to update the SBR is described in Box 6.14 below.

1.32.   Within a national statistical office, the SBR should be the single, central place where statistical units are derived and maintained for economic statistics. In this way the economic behaviour of various populations of enterprises can be compared over space and in time. The coordinating role of the SBR is even stronger when statisticians gathering or compiling statistics refer to, and use, the units provided by the SBR without further modification. This is contingent, however, on their satisfaction with the quality of the data provided by the SBR and hence with the administrative sources on which these data are based.

1.33.   In many countries other institutions, such as government agencies or the central bank, also compile official business statistics. With regard to SBRs in these cases, the national statistical office faces additional challenges in fulfilling its coordinating role to ensure consistent official statistics. If legally possible, the SBR of the national statistical office should serve as the central register and other institutions needing register data for their surveys should have access to the SBR of the national statistical office. As a minimum, some kind of register cooperation should be pursued, as maintaining different SBRs is not only costly, it may also result in different survey frames and thus lead to non-harmonized business statistics. Cooperation with national central banks could, in particular, enhance the coverage and quality of the SBR in the financial sector.

1.34.   The demand for better, quicker and more detailed business statistics, and the need of policymakers for more comparable international economic data, is likely to force many national statistical offices to extend the roles of their SBRs and create a multipurpose system which supports statisticians in improving the efficiency of their processes and quality of their outputs.

1.4 Aims of the Guidelines

1.35.   The main objectives of the Guidelines are to:

(a)   Provide practical guidance on core issues of establishing and maintaining the SBR;

(b)   Clarify typology, concepts and definitions, including for statistical units;

(c)   Provide guidance on the use of administrative and other sources for the establishment and updating of the SBR;

(d)   Provide guidance on how to use the SBR in its own right for production of statistics and how information from the SBR can be combined with information from other statistical registers, administrative sources or surveys to produce new statistics;

(e)   Provide guidance on the role of SBRs in the modernization of statistical production and services;

(f)   Provide guidance on issues that are of particular relevance for developing countries.

1.36.   These Guidelines are extensively based on the 2015 Economic Commission for Europe (ECE) Guidelines for Statistical Business Registers. As requested by the Statistical Commission in 2015, the present Guidelines expand the ECE Guidelines with examples from countries from all regions and with practical guidance that is applicable to a broad range of statistical systems. Although the recommendations in these Guidelines may be viewed as aspirational in some countries, they should be implemented according to national circumstances and they should guide the continuous process of improvement of the SBR.

1.37.   The Guidelines are targeted at SBR management, SBR staff members, the staff of business statistics, and the staff dealing with respondent relations and with administrative authorities that deliver data to the SBR. They may also be useful for training purposes.

1.38.   These Guidelines are not sufficiently detailed and comprehensive to cover all the SBR issues, concepts, definitions and methods that are important in every country. They do not and could not take full account of all the various national institutional structures and concerns. Like any other international manual, the Guidelines can only provide guidance in the form of a broad range of concepts and explanations that need to be interpreted within each particular national context, which will invariably differ from country to country.

1.39.   The main conceptual framework on which the Guidelines are based is the 2008 SNA, which is the overarching framework for integrated economic statistics and therefore the natural reference for the concepts and definitions of the relevant terms of the SBR. International classification manuals, in particular the fourth revision of the International Standard Industrial Classification of All Economic Activities (ISIC Rev. 4), are also referenced.

1.40.   A variety of other international guidance manuals that deal with specific phenomena or domains should also be mentioned. They include manuals on the use of administrative data, on the informal sector, on business demography, on foreign affiliate statistics, on environmental-economic accounting, on integrated economic statistics, which also addresses the key role of the SBR, and on measurement of global production.

1.41.   All these international manuals and various other documents, papers and reports were consulted and used in developing and drafting the present Guidelines

and the ECE Guidelines. There are, however, two sets of international guidelines which directly focus on the SBR: the European business statistics methodological manual for SBRs of Eurostat (2021) and the Guidelines for Building Statistical Business Registers in Africa (African Development Bank, 2014). These two manuals have been extensively used in line with the principle applied in developing the Guidelines that already existing material should be used whenever feasible and appropriate.

1.42.   As already stated, international manuals and guidelines can provide concepts and definitions needed for harmonized statistics, and thus help in developing national implementation. The exchange of experience between countries and bilateral cooperation arrangements are also recommended, however. There are two main international forums for the exchange and discussion of national experience, country practices and international developments in the area of SBR. One is the Wiesbaden Group on Business Registers – a city group under the auspices of the Statistical Commission, which organizes biennial meetings – and the second is the biennial, joint ECE, Eurostat, and OECD Expert Group Meeting on Business Registers, which takes place in years where there is no Wiesbaden Group meeting. In addition, the task teams on exhaustive business registers and on capacity-building of the United Nations Committee of Experts on Business and Trade Statistics hold regular meetings to discuss and advance the concepts and best practices in the development of SBRs.

1.43.   One of the core goals of the update of the Guidelines was to focus on the problems and issues of countries with less developed statistical systems. Accordingly, additional issues and country examples have been added to illustrate the specific situation and to show how countries deal with and overcome these issues. Nevertheless, the situation in each country is different and a single solution is not feasible. For that reason, the overall recommendation is that each country draft a strategy for the creation and development of their SBR in line with the principles outlined in the present Guidelines.

1.44.   The availability of administrative data sources for the update and maintenance of the SBRs is often a major problem in developing countries. It is therefore recommended that countries establish a strategy for the development and access of data sources. In box 1.4 the main elements of such a strategy are outlined.

Box 1.4
Recommendations for a strategy on data sources

In countries where there is a lack of solid administrative registers, it is necessary to take advantage of different sources for constructing and updating the SBR; for example, to use information provided by several surveys, censuses, and available registers about specific sectors (such as data provided by certain chambers) with the aim of combining these sources and incorporating at least a list of establishments, enterprises and enterprise groups into the SBR.

To this end, it is recommended that a clear strategy be defined with the necessary steps in order to make optimal decisions for updating the SBR using all the available information. For this purpose, national statistical offices should be aware of the following needs:

➤ To take advantage of all the information collected by the national statistical offices through activities that involve field operations with the purpose of ensuring the identification, location and characterization of businesses.

➤ To identify public and private sources of administrative registers that contain data on businesses.

➤ To select an economic activity or group of establishments or enterprises for which the update is easier, more practical and maintains the reception of robust and periodic information, regardless whether it originates from the outcomes of field operations or administrative sources.

➤ To adopt and promote the use of a unique identifier. One of the most relevant challenges is the linkage among different sources, so it is necessary to ensure that the linked information refers to the same economic unit by assigning a unique identifier to facilitate its follow-up in the SBR.

➤ To use the tax register as a benchmark in order to determine the coverage of the SBR and the advantages that it could bring in terms of the previous need.

➤ The national statistical offices should define a long-term plan for strengthening the use of administrative registers in the updating of the SBR; for this purpose, the following actions are recommended:

• Examine the regulatory framework for the production of economic statistics in the country, in order to ensure the sharing the administrative registers held by government agencies and private organizations.

• Prepare inter-institutional agreements that facilitate information exchange for statistical purposes.

• Assess and improve the quality of administrative registers to contribute to strengthening the information in the SBR.

• Promote the use of a unique identifier for each business in the different sources.

1.5 Overview of the Guidelines

1.45.   The Guidelines are divided into 12 chapters and 5 annexes. Each chapter may be read separately, according to the interests and needs of the reader. Cross-references are provided. In the following paragraphs, a short description of the topics and contents of each chapter and of the annexes are provided.

1.46.   Chapter 2 on the roles of the SBR provides an overview of the various roles that SBRs play in the context of the production of business statistics. In total, eight differ-

ent roles are distinguished and described. The first role is the central role, namely the maintenance of a set of statistical units by providing a gateway between data from various input sources (mainly administrative sources) and statistical units. This results in continuous updating of the content of a *live register.* The second role is to provide *register snapshots* and *frozen frames*, from which *survey frames* are derived. These roles are the most central and important ones and focus on the support function of the SBR for business surveys. An additional role is direct support for the surveys by survey registration, monitoring survey response and measurement and control of response burden.

1.47.   Chapter 2 also describes additional new roles that have emerged in the last years and show the usefulness of SBR databases. The most important is the production of statistics directly based on the SBR. Examples are economic censuses and business demography statistics. In some countries, individual data on enterprises can be made public and thus support directories or lists of businesses or georeferenced information systems. The processes of globalization have rendered the international comparability of statistical data even more important. A key enabler would therefore be the exchange of microdata between countries through the SBR, which would significantly enhance the quality of global statistics. A related SBR role is the integration of the SBR in the production process of economic statistics.

1.48.   Chapter 3 describes coverage of an SBR, examining the concept from the standpoint of various institutional unit types and sectors of the economy as defined in the 2008 SNA. This chapter demonstrates how, in principle, the SBR should record all units in the national economy that contribute to gross domestic product (GDP) but observes that, in practice, this is not achievable because of various constraints, which are described. The chapter deals with the relevant concepts, such as market and non-market producers, the observed and non-observed economies, the informal sector and illegal activities. It provides recommendations for the minimum coverage of an SBR.

1.49.   Chapter 4 and chapter 5 deal with the concepts and characteristics of statistical units. While chapter 4 introduces the various statistical units and their delineation, chapter 5 provides information on their characteristics, which should be recorded in the SBR. The main statistical units discussed in chapter 4 are the *enterprise group,* the *enterprise,* the *establishment,* the *kind-of-activity unit* (KAU) and the *local unit.* The chapter also deals with various forms of legal and administrative units that constitute the main building blocks for delineation of the statistical units. The chapter also provides guidance on specific kinds of units, such as holding companies and other special purpose entities (SPEs), and also on units for specific economic sectors, such as agriculture and government.

1.50.   Annex A and annex B are connected with chapter 5. Annex A lists and explains in detail the characteristics of the legal/administrative units and the various statistical units. Annex B introduces the international activity classification ISIC Rev. 4, and the SNA classification of institutional sectors, which are the two most important statistical classification systems used in the SBR.

1.51.   Chapter 6 deals with the data sources for the establishment and maintenance of the SBR. It considers not only the various administrative data sources, which are the main sources of input information, but also all other sources. The chapter describes the general methods, procedures and issues in relation to the data sources, in particular cooperation with the administrative authorities. It gives advice on the use of administrative data and the problems that may occur when administrative data are used for statistical purposes. The chapter also gives an introduction to the record linkage methods that are used when the linking of individual units in the various data

sources is not supported by common identifiers. Methods based on other data, such as name and address, must be applied to identify the same units in the different sources so as to link them correctly.

1.52.   Related to chapter 6 is annex D3, which provides an excerpt from the administrative data source evaluation checklist developed by CBS.

1.53.   The concepts of the update and maintenance of the SBR are the subject of chapter 7. This provides the fundamentals and basic considerations for the establishment of a maintenance strategy that is based on the needs of the users, in particular survey staff. The chapter deals with the recording of demographic events, such as enterprise births, the handling of the changes in the characteristics, the continuity rules and, last but not least, the treatment of errors. Annex F provides examples of the methodology and practices followed in profiling (Eurostat, Canada, Colombia, Indonesia, Netherlands, South Africa, Switzerland and United Kingdom of Great Britain and Northern Ireland).

1.54.   Chapter 8 covers survey frame methodology. Starting from the Generic Statistical Business Process Model (GSBPM), the specific roles of the SBR are described in relation to the survey requirements. The various population concepts lead to a discussion of frame design. The chapter deals with the effects of overcoverage and undercoverage errors and their correction, along with other kinds of frame errors.

1.55.   The emerging role of producing statistics directly based on the SBR is discussed in chapter 9. As previously noted, the SBR may itself now also have a dissemination role, whereas in the past only survey results were published. The full coverage and good timeliness of SBR data are used to produce census-type statistics and business demography statistics. Dissemination also requires the consideration of confidentiality rules, which differ between countries. The provision of microdata requires a specific legal basis. In a few countries, it is possible to provide microdata to researchers or even to publish individual business information, while in most countries this is not possible. Annex E1 gives an overview of the Italian register-based census as an example of statistics directly based on SBR.

1.56.   Chapter 10 discusses SBR quality. As already mentioned, SBR quality is a crucial issue as it has a major impact on the quality of business statistics, especially in terms of completeness of coverage and accuracy of the characteristics (such as economic activity code) used for survey frames and sampling. Beginning with a definition of quality and an explanation of the quality dimensions related to the SBR, the chapter deals with the methods of quality assessment. It proposes quality indicators and provides the basic elements of a quality improvement policy. Annex D1 describes the Italian SBR quality indicators and annex D2 the Colombian experience in implementing quality processes.

1.57.   Chapter 11 on key considerations in establishing an SBR is of particular interest for countries where an SBR is being developed or under review. It deals with three related areas: planning, governance and organizational and IT considerations. It covers the various aspects that need to be taken into account in the process of establishing an SBR (including long-term scoping and the modular approach). It deals with the legislative framework, funding and human resources, and discusses and provides guidance on IT topics such as system architecture, database management, programming and software tools. A particular case of SBR architecture from Statistics Canada is described in annex C3. Annex E2 illustrates the calculation of a check digit for an identification number.

1.58.   Chapter 12, the last chapter, lists and briefly describes topics for further work and research as recommended by the Task Force. These topics were identified during

the course of development of the Guidelines and are prompted by the need to deal with emerging issues that will pose major challenges in the future. As the development of new methods and good practices is resource-demanding, experience should be shared among countries and common development projects should be encouraged. The topics identified include the concepts and delineation of statistical units, the use of administrative data sources, the exploration of new data sources, the challenges of economic globalization, the role of SBR as the backbone for economic statistics, and the role of SBR in the modernization of statistical production and services.

1.59. Annex C presents examples of SBRs in Canada, Costa Rica, Denmark, Georgia, Indonesia and Malaysia.

1.60. The abbreviations and acronyms used in the Guidelines are listed at the beginning, before the table of contents. At the end of the Guidelines, there is a glossary, listing the key terms used in the Guidelines with brief definitions and explanations, and a list of sources cited in the text.

1.6 Terminology

Introductory remarks

1.61. As in guidelines of any sort, it is vital that all users have the same understanding of the key terms used. The following paragraphs identify and define the key terms used in the present Guidelines and, throughout the text, an endeavour has been made to avoid the use of other terms that are synonyms but could be interpreted as having slightly different meanings. Where a term has synonyms in wide use, the approach in these Guidelines has been to use only one of those terms and to reference its synonyms the first time that this selected term is used. In the cases where the terms are not exact synonyms, the intended distinctions between them are made clear the first time they are used.

1.62. The definitions in the document and summarized below are based on terminology defined in the 2008 SNA and ISIC Rev. 4, supplemented where necessary by terms from the Guidelines for Building Statistical Business Registers in Africa developed by the African Development Bank. Additional information is provided in the Glossary.

1.63. Such terms as *business, company, establishment, enterprise, unit, statistical unit, respondent, economic unit, economic organization, economic operator, economic producer, legal entity, legal unit, local unit, legal local unit* and others may have different meanings in the literature. In the Guidelines they have the following specific meanings:

> ▸ *Unit* – a single distinct part or object, as generally understood in the English language, needing further qualification to have a more precise meaning.

> ▸ *Entity* – a synonym for unit.

> ▸ *Legal unit* – a unit that is recognized by law or society, independently of the persons or institutions that own it. The characteristics of a legal unit are that it owns goods or assets, it incurs liabilities and it enters into contracts. A legal unit always forms, either by itself or sometimes in combination with other legal units, the legal basis for a statistical unit.

> ▸ *Economic unit* – a legal unit, or part of a legal unit, with economic production as defined in the 2008 SNA.

‣ *Economic operator, economic producer,* and *economic organization* are synonyms for economic unit, not used except where they appear in a quotation.

‣ *Administrative unit* – a unit defined by a legal unit for the purposes of conforming with an administrative regulation, for example VAT.

‣ *Operational unit* – a unit defined by a legal unit for the purposes of organizing itself, for example a division, branch, workshop, warehouse or outlet.

Standard statistical units

‣ *Statistical unit* – a unit defined for statistical purposes; the basic unit of observation within a statistical survey; the unit for which information is sought and for which statistics are ultimately compiled. There are four types of statistical unit defined in the 2008 SNA and ISIC Rev. 4, as defined below. In addition, enterprises may be grouped by legal or financial links or both into an enterprise group.

‣ *Enterprise* – a legal unit (or the smallest set of legal units) that produces goods or services and that has autonomy in respect of financial and investment decision-making. An enterprise may be a corporation (or quasi-corporation), an NPI, or an unincorporated enterprise. An example of an unincorporated enterprise is a household or government unit in its capacity as a producer of goods or services.

‣ *Establishment* – an enterprise or part of an enterprise that is situated in a single location and in which only a single (non-ancillary) productive activity is carried out or in which the principal productive activity accounts for most of the value added.

‣ *Local kind-of-activity unit (LKAU)* – synonym for establishment.

‣ *Kind-of-activity unit* – an enterprise or part of an enterprise that engages in only one kind of productive activity or in which the principal productive activity accounts for most of the value added.

‣ *Local unit* – an enterprise or a part of an enterprise (e.g. a workshop, factory, warehouse, office, mine or depot) that engages in productive activity at or from one location.

‣ *Institutional unit* – an economic entity that is capable, in its own right, of owning assets, incurring liabilities and engaging in economic activities and in transaction with other entities. The term covers both market as well as non-market units.

Other types of units

‣ *Business* – (loosely used) a type of enterprise, such as a commercial enterprise or legal entity engaged in a commercial economic activity.

‣ *Corporation* - legal unit created for the purpose of producing goods or services for the market that may be a source of profit or gain to its owner; collectively owned by the shareholders who have the authority to appoint directors responsible for its general management.

‣ *Company* – synonym for corporation.

> ▸ *Firm* – used to refer to a company, partnership or unincorporated sole proprietorship.

> ▸ *Observation unit* – a unit about which data are obtained during the course of a survey; usually a statistical unit, or, if data cannot be obtained about a target statistical unit, then some other unit about which data can be obtained and from which data for a statistical unit can be compiled.

> ▸ *Reporting unit* – the unit from which data about an observation unit are obtained during the course of a survey. It may or may not be the same as the observation unit. An example of it not being the same is where an accounting business reports data on behalf of a client business that is the actual subject of the survey.

Data elements associated with SBRs

> ▸ *Register* – as generally understood in the English language, a database on items or events, often kept by the official whose job it is to do so; the term requires further qualification to have a precise meaning.

> ▸ *Business register* – database on businesses, or more generally on any administrative or statistical units, kept for commercial, administrative or statistical purposes, including relevant characteristics of the businesses or units.

> ▸ *SBR* – database on statistical and other units kept by a national statistical office for statistical purposes.

> ▸ *Live register* – the part of an SBR that is being continually updated with new information about the units and their characteristics.

> ▸ *Register snapshot* – a copy of the live register at a given point in time.

> ▸ *Register picture* – synonym of register snapshot.

> ▸ *Frozen register* – synonym of register snapshot.

> ▸ *Frozen frame* – register snapshot containing only active statistical units.

> ▸ *Common frame* - synonym of frozen frame.

> ▸ *Frame* – set of statistical units, which forms the actual population for a survey.

> ▸ *Frame population* – synonym of frame.

> ▸ *Historical register* – historical versions of the live register at points in time in the past; can be through a set of consecutive register snapshots.

Miscellaneous other terms

1.64.   The terms *enterprise* and *economic statistics* are used in preference to the narrower terms *business* and *business statistics* as an SBR may well include lists of government units and non-profit organizations, which are enterprises but not businesses, and which are the subject of some surveys (for example employment) for which frames are derived from the SBR.

1.65.   The terms *characteristic, variable, property* and *attribute* as applied to units in the SBR are considered as synonyms and the term *characteristic* is used in the context of the SBR and survey frames. The term *variable* is more appropriate in the case of surveys where sampling is involved, introducing the notion of variability due to the probability mechanism involved in selecting the samples.

1.66.   Any type of survey conducted by SBR staff specifically to improve SBR quality is referred to as an *SBR improvement survey*. Synonyms include quality improvement survey, nature-of-business survey, control survey and SBR survey, SBR coverage survey and SBR quality improvement survey.

1.67.   With regard to the organizations or persons using SBR outputs, the term *user* is used, rather than *recipient, receiver* or *customer*. Most SBR users are internal to the national statistical institute, being the staff of surveys that draw frames from the SBR.

1.68.   In the context of surveys (including SBR improvement surveys) the term *response burden* is used in preference to synonyms such as *respondent burden, reporting burden* and *administrative burden*.

Box 1.5
Examples of strategies followed by countries in developing their statistical business registers

The establishment of a comprehensive SBR is usually a long-term process and needs to be organized in a stepwise procedure. The main criteria for a development strategy are the given user demands, the available human and technical resources and the data sources that are either available or need to be collected for the maintenance of the SBR. Further details on this may be found in chapter 11 on the key considerations in establishing an SBR.

Examples illustrating this process in different countries and describing their strategy in the development of their SBRs are provided below.

Colombia

In Colombia, the National Administrative Department of Statistics has established a programme to strengthen the country's SBR, with the aim of improving the quality of the statistics of public and private entity registers and promoting their use in the generation of official statistics, in order to expand the statistical information provided to the country. The programme has three components: diagnosis, design, and anonymization of administrative registers. Each of these is documented in terms of the activities to be developed and the tools designed for each case. (further information is available at www.dane.gov.co/index.php/sistema-estadistico-nacional-sen/registros-administrativos/programa-de-fortalecimiento#l%C3%ADnea-2). At the diagnostic level, a methodology has been designed and disseminated for administrative registers for statistical use (available at www.dane.gov.co/files/sen/registros-administrativos/Metodologia-de-Diagnostico.pdf). This methodology allows for an assessment of the data quality of administrative registers to determine their fitness for use as data sources for statistical business registers, based on minimum technical characteristics .

Egypt

In Egypt the main problem in building the country's SBR is that the statistical law is not fully in place: the tax authorities have reservations about sharing their database with the main statistics authority, the Central Agency for Public Mobilization and Statistics (CAPMAS), as they feel that by sharing the data they will forfeit their confidentiality.

In addition, data from some other administrative sources, like the Social Security Authority, are of low quality. Field tests have shown that some establishments do not exist in the field, while others have closed down or changed their activity.

CAPMAS has a memorandum of understanding with the Social Security Authority and has access to its database, which is updated every three months. This database contains several items, namely: establishment name, insurance number, address, economic activity, legal entity, number of the employees who have insurance, the commercial register number and the activity start-up date.

Box 1.5 (*continued*)

In Egypt there is now an economic census which can provide the following information:

- Business name of the establishment
- Number of employees
- Main and secondary economic activity (up to three secondary economic activities)
- Address details and contact details
- Structure of the enterprise groups, such as the address of the headquarters and its branches
- Legal entity
- Regulatory laws of the establishments
- Value of the paid capital
- Value of the invested capital
- Exact date of incorporation and the activity start-up date
- Exact working months in the year
- Number of working employees and their salaries
- Value of the industrial production

All this information will be useful in building the SBR together with the data from Social Security Authority.

State of Palestine

Over the past four years, the focus was on creating the basic requirements for building and developing the SBR, starting with constructing the infrastructure facilitating the use of administrative registers rather than censuses as a basic source of the SBR, even if their coverage only extended to large and medium enterprises as a first stage. The procedure is the following:

- Studying and describing the contents of the administrative registers used for licensing in accordance with the law, in terms of their coverage (comprehensiveness) and information (variables and quality).
- Setting the minimum required variables with the administrative partners to ensure the unification of the common identification number. Where the commercial name of the establishment and the licence number were specified (the company ID is officially registered by the Ministry of National Economic Affairs, but there is no unified ID for the other companies), they include: the name(s) of the owner(s), the ID of the owner(s), address and phone numbers.
- Linking the data of the administrative registers with the 2017 census data and constructing a database of intersecting establishments between all sources, entitled Version 0 of the Administrative Business Register (ABR (V0)).
- Developing interchange mechanisms among all partners in the establishment registration system.
- Providing a legal framework (cabinet decision) for the adoption of variables, IDs and interchange mechanisms among all partners to ensure that the SBR is updated regularly.
- Enhancing the quality and the coverage of the ABR (V0) by conducting a field listing of all the large establishments (with 10 and more employees) in the census that were not matched with at least one administrative register with a view to improving and ensuring the quality of the identification information and thereby increasing the chance of a match (ABR (V1)).
- Continuing the development of ABR to increase its coverage and quality in parallel with the usage of the establishment census as a basic source of conducting the framework under the low coverage of the register.

Chapter 2
Roles of the statistical business register

2.1 Introduction

2.1.   As discussed in chapter 1, an SBR is a vital component of the core statistical infrastructure supporting the collection of economic data and production of economic statistics. The SBR serves as the coordinating mechanism for economic statistics and ideally provides all the information needed by the statisticians responsible for the various economic surveys and related statistical processes and outputs. The SBR may therefore be said to play the role of a backbone in the production of economic statistics.

2.2.   This chapter presents the specific roles of the SBR, which are listed in figure 2.1 below. Each role is described in detail, indicating the inputs needed to perform the role and the statistical outputs produced. The chapter first presents the primary roles of an SBR in maintaining and providing sets of statistical units, in particular enterprises, and in providing frames and other supports for surveys. It then lists a number of other SBR roles that are related to the primary purpose and build upon the methodology and databases of SBR. The chapter draws on material from Eurostat working groups and international conferences and meetings and includes examples from the national statistical offices in Italy, Mexico, the Netherlands and South Africa.

2.3.   The first role *(SBR live register)* of the SBR listed in figure 2.1 is to maintain a set of statistical units as the foundation for creating frames for economic surveys. The SBR is updated on a daily basis, with new information from various sources. This information is used to create and deactivate statistical units, and to update their characteristics. This functionality of continuous creating, updating and deactivating of the corresponding units is referred to as the "*live register*".

Figure 2.1
Overview of the SBR roles

	Role	Purpose
1	SBR live register	Provide the gateway between data from various input sources and statistical units
2	SBR register snapshot and frozen frame	Provide populations of statistical (and possibly administrative) units at fixed points in time
3	SBR survey frame	Provide a set of statistical units for a survey, valid for a specified reference period, with all the characteristics required
4	SBR survey support	Monitor survey response and measure and control response burden
5	SBR statistics	Produce statistics based directly on the SBR
6	SBR information source	Provide lists of enterprises and their locations and possibly other characteristics
7	SBR international data exchange	Facilitate coherence in international statistics
8	SBR in modernization of statistical production and services	Promote integration of the SBR within the production processes for economic statistics

2.4.   The second role *(SBR register snapshot and frozen frame)* is to produce register snapshots and frozen frames from the live register as the basis for subsequent generation of survey frames. A *register snapshot* contains the set of statistical units valid for a specific reference period and hence represents a coordinated population of statistical units in space and time. A *frozen frame* contains all the active statistical units in a snapshot that have sufficient information associated with them for sample selection purposes. A snapshot (or frozen frame) may also provide the links between statistical and administrative units (for data collection) for the specific reference period.

2.5.   The third role *(SBR survey frame)* of the SBR is to provide frames for surveys. For each survey the SBR provides the set of statistical units in scope for the survey (typically a subset of the frozen frame) together with the characteristics required for stratification and sampling, and, for the selected units, the information required to contact and communicate with them.

2.6.   The fourth role *(SBR survey support)* concerns survey registration and survey control. It involves tracking the reporting statuses of enterprises and the response burden on enterprises imposed by surveys. The SBR supports this by storing information that can be used to optimize survey design, as well as to monitor reporting statuses and to compile overall response burden.

2.7.   The fifth *(SBR statistics)* and sixth *(SBR information source)* roles are to provide macrodata (economic and demographic statistics), and microdata (enterprise names, address and characteristics), respectively, to users external to the national statistical office. This also includes data provided for analytical and economic research purposes. Such dissemination is constrained by the confidentiality provisions applying to the national statistical office, in particular in the case of microdata.

2.8.   The seventh role *(SBR international data exchange)* is to facilitate exchange of data and metadata across countries in order to improving coherence and international comparability of economic statistics. This is particularly important to ensure coherence of cross-border economic statistics, such as inward and outward FATS between two countries, and for the measurement of activities of multinational enterprises.

2.9.   The eighth role *(SBR in modernization of statistical production and services)* refers to the role of the SBR in coordinating, linking and connecting units that occur in all kinds of sources relevant for statistics. The SBR facilitates the use of administrative registers and is a potential gateway for combining data from traditional sources and new media, and may serve as a data warehouse providing integrated data to users. It can also facilitate the new, industrialized and standardized production of statistics as part of a process of modernizing the statistical production and services, at both national and international levels.

2.2   SBR live register

2.10.   Statistical units in economic statistics are used to describe in a standard way the various populations of organizations or parts or groups of organizations involved in economic production. An important role of the SBR is to maintain, and to keep track of changes in, statistical units and their characteristics that occur in the economy. Maintenance is a continuous process in which constant modifications of the set of statistical units occur over time. The extent of the modifications depends on the update strategy of the SBR. In this respect the SBR is considered to be a *live register* in which the composition and characteristics of units continuously change over time.

2.11.   The live register is a vehicle for bringing together data from the various sources that provide the basis for derivation of statistical units. It is the starting point for communications with the owners of the sources. Legal units are usually the building blocks for creating statistical units. In some countries the SBR is the only environment in which legal units of all forms are brought together.

2.12.   Statistical units are created in the live register. There are several types of statistical units. The most important of these is the *enterprise*, usually defined in accordance with an international definition (as further discussed in chapter 4), although implementation may require some compromise.

2.13.   An economic census used to be one of the main sources from which to derive a common frame in order to conduct and coordinate business surveys. Nowadays SBRs are usually based on legal and administrative units obtained from a network of administrative registers, SBR improvement surveys, and feedback from economic surveys, as illustrated in figure 2.2.

2.14.   The processes for the creation and maintenance of statistical units, in particular the set of enterprises, are the most important factors in determining the quality and the usability of SBR populations for economic statistics. A well-defined maintenance strategy is a key aspect of an SBR. The units in an SBR are maintained and updated with the most recent information available, as further discussed in chapters 6 and 7.

2.15.   Timeliness and linkage issues can pose problems when using multiple sources, but their use in combination enables a national statistical office to derive all relevant statistical information and hence to reduce response burden and create new and more detailed outputs.

Figure 2.2
SBR live register
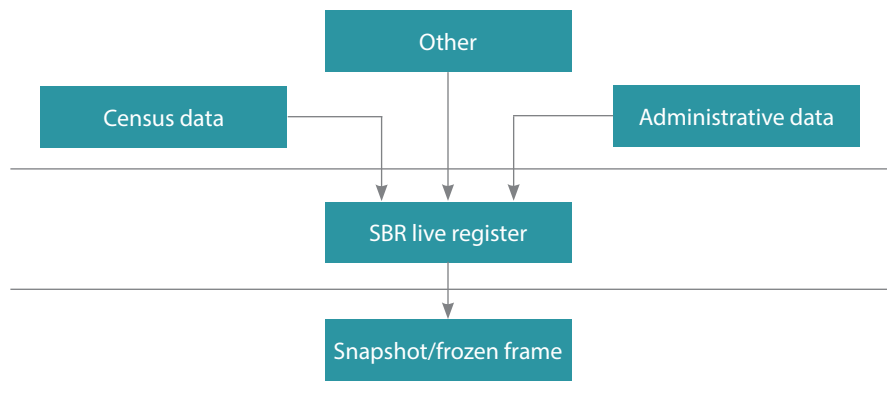
2.16.   In order to make full use of source information, the values of key characteristics of units in the sources should be harmonized using well-defined standards, for example, for dates, telephone numbers, addresses, legal form, economic activity, number of employees, and other characteristics, as further discussed in chapter 5.

2.17.   Monitoring the continuity of statistical units is also an important aspect of SBR maintenance. In the event of new information about a statistical unit, a decision has to be made whether the unit keeps its identity in the SBR, is deactivated in the SBR, or is registered as a new unit. This is further discussed in chapter 7.

2.18.   When updating administrative and statistical units and their characteristics, the existing and previous values should be retained in the live register if possible, or at least be retained in a *historical register*. This enables any change over time in any unit used for statistical production to be traceable back to the source, which may help understanding anomalies in statistical processes. Key SBR users, particularly those with direct access to the live register, must be trained so that they are aware why historical versions exist and why they may differ for certain units for different extraction dates.

2.19.   Units from different sources are linked at micro level. However, not all legal or administrative units lead to the creation of a statistical unit or are necessarily linked to a corresponding statistical unit. This can happen because a legal or administrative unit is dormant, or inactive, or the information to determine whether the unit is active is simply not available in the source data. Thus, some legal or administrative units are not represented in the frozen frame since it includes only statistical units and are not, therefore, used in constructing survey frames or in subsequent compilation of statistical aggregates. These *floating* legal and administrative units and their data should, however, remain available in the live register. Survey statisticians may need the information about such units to impute missing data, or to conduct surveys on specific groups of missing units, or to help in profiling.

2.3   SBR snapshot and frozen frame

2.20.   The most important role of an SBR is its coordinating task in the production of economic statistics. This backbone role is best fulfilled if all statisticians in a national statistical office use the units delineated in the SBR as of specified reference dates. In other words, it must be possible for statisticians to view and retrieve a common set of units from the SBR as of a specific reference date.

2.21.   Thus, while the live register continuously changes, statisticians need coordinated input for their processes. If surveys are conducted with different time lags from a given reference period, the composition and the quality of the population of statistical units in the live register at the times the survey frames are drawn will differ. This creates the need to derive a "*frozen version*" of the population of statistical units – a composition of units that is valid for a particular moment in time for use by all surveys. This is realized by taking a "*snapshot*" of the live register (which itself may be viewed as a *continuously changing film*) as illustrated in figure 2.3.

2.22.   Each *snapshot* is extracted from the live register at a point in time. The set of all snapshots represents a set of coordinated populations of statistical and administrative units at specified time points, with a certain quality level. More precisely, a snapshot at date t consists of:

- ➤   Statistical units at t
- ➤   Administrative units linked to statistical units at t (or *links* to these administrative Units), and possibly
- ➤   Administrative units not linked to statistical units at t, but in the live register at t

Figure 2.3
SBR frozen frame

SBR live register

Register snapshot frozen frame

Set of frozen frames
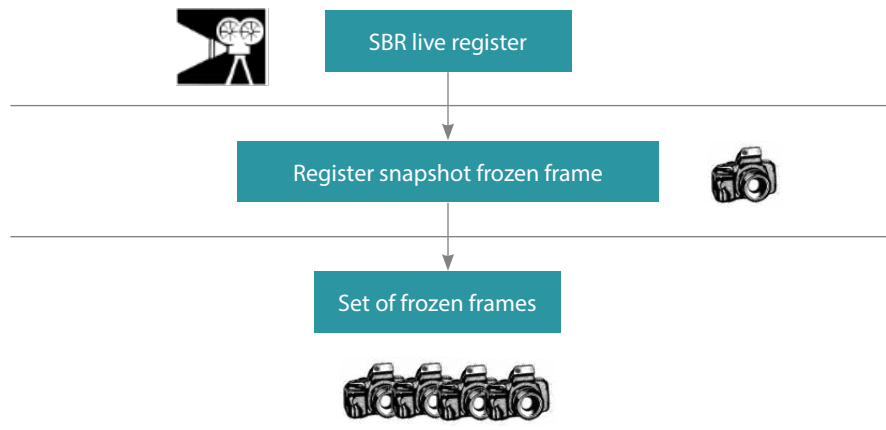
2.23.   The units in a particular register snapshot that are "*active*" as of the specified reference period together with all the information about their characteristics needed for survey sample selection and other statistical purposes, constitute the "*frozen frame*" for that period. More specifically, a frozen frame at date *t* consists of:

▸   Statistical units active at *t*, and with sufficient information for survey sample selection, and

▸   Administrative units linked to statistical units at *t*, or links to these units.

2.24.   In some literature on this topic a frozen frame is called a *master frame*, or a *common frame*. This is because typically (as discussed below) several survey frames are derived from this single common extraction from the live register, thereby coordinating the survey frames at a moment in time.

2.25.   The quality of a frozen frame depends on the quality of the information about the existence of units and their characteristics and on the reference period for which statistics are being compiled. This should be taken into account in trying to ensure that short-term and annual indicators, which refer to overlapping reference periods, are coherent. It may lead to the revision of previously published short-term indicators when annual indicators are produced. If a statistician needs more information about certain units than is available from the frozen frame, then this extra information may be obtained by linking back to the same units in the live register.

2.26.   In summary, a snapshot is a copy of all the statistical units in the live register, and all administrative units or at least all links from statistical units to administrative units. It is an intermediary step between live register and frozen frame. It is used to check for errors that have crept in during processing since the previous snapshot, also as the basis for a historical record (as further discussed below). It may contain many inactive enterprises. A frozen frame is a subset of the snapshot that comprises all statistical units that are active, or potentially active, or active within the previous reference year. It also includes administrative units that are linked to these statistical units. Its aim is to

include all units and all characteristics that are used by subsequent processes and nothing else. In other words, it is a trimmed down version of the snapshot that is easier to manipulate because the (possibly large number of) inactive units are not there.

2.27.   If the frozen frame for an annual survey for reference period *t* is also used to observe the short-term indicators in year *t*+1, it will not include the information on newly active units for reference year *t*+1. Accordingly, some national statistical offices create new version of the frozen frame at later points in time for subannual surveys. Thus, frozen frames may be produced annually, quarterly, or monthly. In all cases, it is important to control their use so that the resulting survey frames are as coherent as possible.

2.28.   A set of frozen frames effectively constitutes a *historical register* which:

> ‣   Coordinates populations of statistical units.
>
> ‣   Provides links between statistical and administrative units.
>
> ‣   Links the SBR to other sources (administrative registrations, survey samples) used for statistical outputs.
>
> ‣   Enables the reconstruction of the history of the units that it contains.

2.29.   Statisticians often compare particular populations of statistical units at different reference points in time and try to describe and explain the differences. The basic data for this sort of analysis is available from the historical register. The evolution of a particular population of enterprises may be presented in the form of enterprise demographic statistics.

Box 2.1
Frozen frames for business statistics in Mexico, Mongolia and the Kingdom of the Netherlands

Mexico

In 2010, the National Institute of Statistics and Geograph (INEGI) created the Statistical Business Register of Mexico (RENEM) from information provided by the economic census carried out in 2009. RENEM data are considered confidential. RENEM covers two types of economic units: establishments and enterprises. For each establishment RENEM includes variables related with its identification, location (including geographical coordinates), industry classification, status (active or inactive), and certain economic variables, such as number of employees and revenues. In addition, it stores historical data of the establishments and enterprises.

RENEM is fully updated every five years using the results of the economic census, and partially updated each year for the largest establishments and enterprises (in terms of revenues) through the monthly economic surveys programme and the annual fieldwork of RENEM to update a specific segment of establishments or enterprises.

Since 2015, through RENEM, INEGI takes snapshots twice a year to get a listing of all the economic units at a specific time point. From these snapshots, it is possible to obtain two products by using only the active economic units: the National Directory of Economic Units (DENUE) and a frozen frame.

DENUE is published on the INEGI website, using the identification, location (including geographical coordinates), industry classification, and size (in terms of the number of employees) variables. Besides being used as a public consultation hub, users can download the various updates of DENUE to conduct their own longitudinal and demographic business studies at a number of aggregation levels: geographical, industrial, according to size, and so on. Moreover, since DENUE is public, it also can be used by independent researchers as a master frame to design their surveys.

Box 2.1 (*continued*)

Frozen frames include all the variables of the snapshot, including economic variables: number of employees and revenues, which are used by the technical area of INEGI for statistical survey design. All the survey frames of INEGI are derived from such frozen frames, which are used once each year to coordinate and conduct the national economic surveys.

Mongolia

The National Statistics Office of Mongolia created its first SBR from the result of the 1998 establishment census. Since the first establishment of the SBR database, frozen frames are stored quarterly from the live database.

Frozen frames are used for the quarterly dissemination of statistical units and statistics to the public, organization of enterprise-based sample survey, and cross-checking with the databases of other government organizations.

The frozen frames are updated on a regular basis from official data sources such as the results of the establishment census, which is conducted every five years as stipulated in the Mongolian Statistics Act, the quarterly statistical report of the SBR, and the annual, quarterly and monthly reports of economic statistics, and also from administrative data.

The frozen frames include the indicators (enterprises; name, unique identification number, location, address, activity status, types of economic activity, number of workers, income from the activity, asset value) of statistical units (establishment and legal unit level or administrative units) retrieved from the taxation office.

Kingdom of the Netherlands

The Netherlands SBR is the repository of statistical units and of administrative units taken from the country's Trade Register and Tax Office. This information is needed to describe the financial and production processes of resident enterprises.

Statisticians within CBS are obliged to use the SBR in producing economic indicators based on coordinated enterprise populations. Every month, a current frozen frame is derived from the live register. This frozen frame represents the coordinated population valid for that month, which comprises enterprise groups, enterprises and local units, and their main characteristics. In addition, the statistical units are linked to the legal and the administrative units valid for that month. In addition, a list with frame errors referring to activity codes or size classes of statistical units is kept in order to be able to generate several instances of each frozen frame with improved quality levels.

Each month the frozen frame is stored separately and can be consulted for statistical production. This environment includes all past frozen frames and coordinates the populations used for statistical purposes. From this set of frozen frames an authorized statistician can retrieve the microdata needed for a statistical activity, for example, linking tax data and survey data to a frozen frame. In addition, each month CBS also extracts a backbone of administrative units that are not linked to a statistical unit in order to support other domains (register snapshot).

2.4   SBR survey frame

2.30.   A survey frame (sometimes called a sampling frame or a survey sampling frame) is a list of units from which a survey sample is selected, together with the characteristics required to draw the sample according to the agreed design (namely, characteristics required for stratification, sample size determination and sample selection).

2.31.   The frame for an economic survey should be a subset of the frozen frame, comprising the set of statistical and (linked) administrative units that match the specification of the survey target population and are active during the survey reference period, together with the characteristics that will be needed for the survey. Thus, for example, a survey of employment will include active units in all (or at least most) industries that are employers, in other words, will exclude units that are non-employers. A survey of manufacturing will include active units that have an ISIC code in the manufacturing group, whether they have employees or not. A survey of capital expenditure will typically include active units in all industries above a certain size. Thus, the frames for different surveys for a given reference period differ from one another but are extracted from the same common set of units, namely the frozen frame.

2.32.   The primary benefits of survey frames that are extracted from a coordinated set of frozen frames are:

> ‣   Harmonization of the surveys
> ‣   Increased potential for integration of survey data
> ‣   Reduction of costs
> ‣   Prevention of double counting
> ‣   More coherence in the resulting statistics

2.33.   These benefits can only be realized when a single live register is maintained and used to derive the frozen frames used for all relevant surveys.

2.34.   It may be that an optimized sample design cannot be applied because some important stratification variables, such as turnover, that correlate strongly with the key indicator to be compiled are missing from the live register and hence from the frozen frame. In this case it should be possible to enrich the statistical units with characteristics available from other (independent) sources and satellite registers in order to complete the survey frame.

2.35.   A sample design which uses a panel population may be very difficult to apply. In this case information on units in historical samples for similar surveys should be made available.

2.36.   Some national statistical offices exclude from the survey enterprises that have already participated in a previous survey, or are taking part in another survey that is still under way, in order to spread the response burden more equitably across the survey population. This is sometimes referred to as giving the enterprises a survey holiday.

Figure 2.4
SBR survey frame
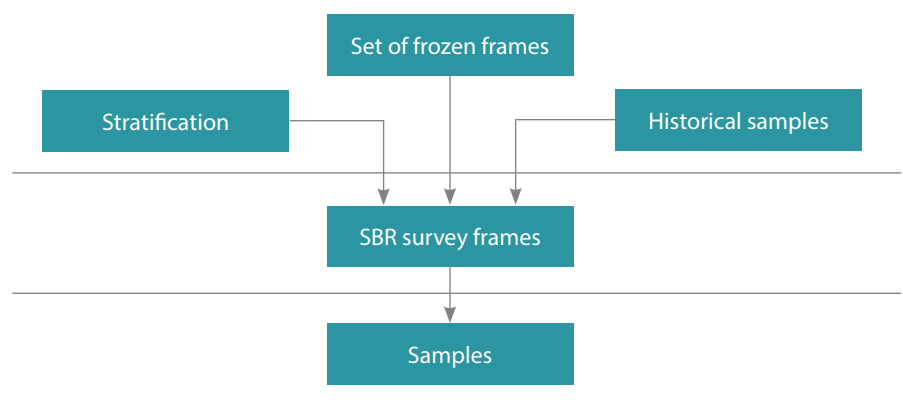
Box 2.2
Survey frames in Brazil, Mexico and South Africa

Brazil

At the Brazilian Institute of Geography and Statistics (IBGE), the SBR is generated every year containing information from administrative data, obtained through agreements with the Ministry of Labour and the tax authority, and the business sampling frame.

Once a year, a frozen frame is provided to the technical department to perform and collect information for the enterprises. IBGE conducts four annual structural surveys, in the domains of trade, service, industry, and construction.

The administrative data provide establishment-level (local units) information, while the structural surveys provide enterprise-level (legal units) information.

Mexico

A frozen frame of the SBR is provided once a year to the technical department to be used as a master frame for the derivation of survey frames. INEGI uses these frames for its regular economic surveys and for specific economic surveys requested from INEGI by other agencies.

The regular economic surveys carried out at INEGI include, among others, at the establishment level: the monthly survey on the manufacturing industry and the monthly survey on services, among others; at the enterprise level: the monthly survey on trading enterprises and the national survey on construction enterprises.

Some examples of requested surveys are the national survey of business financing, at enterprise level and required by the national banking and securities commission; and the survey on information and communication technologies, and the survey on research and technological development, both at enterprise level and requested by the National Council for Science and Technology.

Lastly, the strategy followed by RENEM for the annual review of the largest establishments and enterprises makes it possible to have an updated frame for both the regular economic surveys and the requested surveys, keeping the response rates at a high level.

South Africa

The Statistics South Africa live register has two parts:

- ▸ Administrative part, containing several sets of administrative units maintained from administrative sources.
- ▸ Statistical part (called the business sampling frame), containing statistical units derived from administrative units and maintained from administrative data, survey feedback, profiling and other sources.

The register snapshot contains all statistical units as of the point in time when it is created and includes the values of all the characteristics that may be used for sampling. A snapshot is created, analysed and verified every quarter.

The Statistics South Africa equivalent of a frozen frame is called the common frame. It is created, analysed and verified quarterly, but only used once per year as the basis for the generation of survey frames. It includes all statistical units active with the characteristics (and only these characteristics) that are, or may be, used for sampling. Thus, statistical units that do not have valid values (industry code, size code, provincial code) for these sampling characteristics are excluded. Selected data are disseminated directly from the common frame.

Once per year all survey frames are drawn afresh from the most recent common frame. Contact data for sampled units are subsequently extracted from the live register.

2.5   SBR survey support

2.37.   The role of the survey support is illustrated in figure 2.5. In a typical enterprise survey, information is collected from *reporting units*, which report on behalf of the *observation units* (which are statistical units) in the survey sample (as further discussed in chapter 4). The answers to survey questions are recorded on questionnaires filled out by reporting units directly, or, less frequently, with the aid of an interviewer. To support this process, information on how to contact the reporting units and their relationship to the observation units is required. While this information can be included in the frozen frame and hence in the survey frame, it is more efficiently extracted afresh from the live register for the sampled units only. Exactly what information is required depends on the mode or modes of data collection.

2.38.   Survey support is provided in three areas: data collection, survey registration, and survey control.

Survey data collection

2.39.   The primary activity in conducting a survey is data collection from reporting units. As previously noted, the information needed to contact the reporting unit depends on the mode used for data collection. In some cases, a paper questionnaire is sent to the mailing address of the reporting unit. For a face-to-face interview, a fieldworker has to visit the reporting unit to collect the information, so a physical address is needed. If the information is to be gathered by telephone, a telephone number is required. It is also possible that a reporting unit may submit an electronic questionnaire by email or through a website, in which case an email address is required.

Survey registration

2.40.   Units that are involved in different surveys, or on repeated occasions for the same survey, can be the subject of a policy to reduce or spread the response burden. For example, it may be the policy for any small enterprise to be observed only once every two years or to participate in at most one survey per year. In addition, as previously noted, enterprises that have already participated may be given a *survey holiday*.

2.41.   The survey registration process provides information about the observation units about which data are to be acquired and the reporting units that are to provide these data. The likelihood of contacting a reporting unit is improved when the most recent contact or communication information is available.

2.42.   It is important to provide the reporting unit's preferred response mode for the survey. The response mode may influence the response rate. Making available a range of modes (paper, telephone, face-to-face, web-based or other electronic mode) can improve response rates.

Survey control and response burden measurement

2.43.   To help control the data-collection process, reporting units should be monitored throughout the collection process. When frame errors are detected, they should be recorded. The survey help desk may be the mechanism by which information about errors is received.

2.44.   In order to improve response rates, reminders are often sent. The administration of reminders can be a complicated process, in particular as it depends not only on the official deadline for reporting specified by the survey but also on an estimate of the likely reaction time of the reporting units or of the quality of the answers that they

can provide. All attempts to contact reporting units have to be recorded, not only to ensure that follow-up is efficiently conducted, but to support subsequent enforcement of a response, as further illustrated by the following example from the Netherlands.

2.45.   Reporting unit response rates (by mode in the case of multiple modes) and questionnaire item non-response rates are important in monitoring the quality of the survey outcomes, as discussed in more detail in chapter 10. Sample sizes and reporting unit response rates are also used in calculating target response burden and actual response burden.

Use of an SBR for economic census

2.46.   An economic census is a special type of survey in which the aim is to collect data from all statistical units in the target population. An SBR can be involved in variety of ways.

SBR is not used: The economic census is conducted on the basis of comprehensive area enumeration. The resulting set of statistical units, whether local units or enterprises, may be used subsequently as the basis for creating or updating the SBR. This approach was used in the past but has become less and less common because, first, it requires extensive resources; second, it does not take advantage of existing lists of enterprises and data about them; and, third, it leaves behind a list of units that is impossible to maintain in its entirety because to do so would require an ongoing census.

SBR provides initial frame: In this approach, which might be called a *register-assisted census*, enumerators are provided with an initial list of statistical units (local units or enterprises) based on the SBR, and they correct and supplement this list. This approach makes better use of the SBR but does not assume that the SBR is definitive.

SBR provides the complete frame: In this approach, the census is restricted to units that are in the SBR. Statistical data are collected about some or all of these units. For some units, data may be obtained from administrative sources instead.

SBR provides the complete frame and links to administrative data: There is no statistical data-collection process. The data are obtained entirely from administrative sources through the SBR. This may be termed a register-based census. Further details, including an example from Italy, are provided in annex E1.

Box 2.3
Response enforcement information in the Kingdom of the Netherlands

Many countries have legislation requiring enterprises to respond to questionnaires from the national statistical office. In 2003, the Netherlands introduced a new statistical law. This law states that CBS is obliged to make maximum use of administrative data to derive its statistical output. Another of CBS' obligations is to reduce the response burden on enterprises, for example by utilizing administrative data sources. This implies that the units registered with all these sources should be integrated into the SBR or at least linked to the SBR. In addition, CBS has included fiscal information on the structure of enterprises (available from the finance department) in its SBR. Not all information, however, can be extracted from the administrative sources available. Accordingly, surveys in which data are directly obtained by questionnaire from a relatively small number of respondents are still needed.

By law CBS has the right to require enterprises that do not respond to questionnaires to provide the requested information. In introductory letter that accompanies a survey,

Box 2.3 (*continued*)

the enforcement procedure is explained. If an enterprise does not respond to an official questionnaire sent by the statistical office, then several (predefined) reminders are sent to the responsible contact person. If the enterprise refuses to respond, CBS has the right to fine the enterprise without any intervention of a third part. To assist this process a separate system has been created, containing all relevant SBR information and keeping track of the complete history of all attempts to contact the responsible person in the enterprise. This information can be used as evidence if the enterprise starts a juridical procedure against the statistical office.

Figure 2.5
SBR survey support
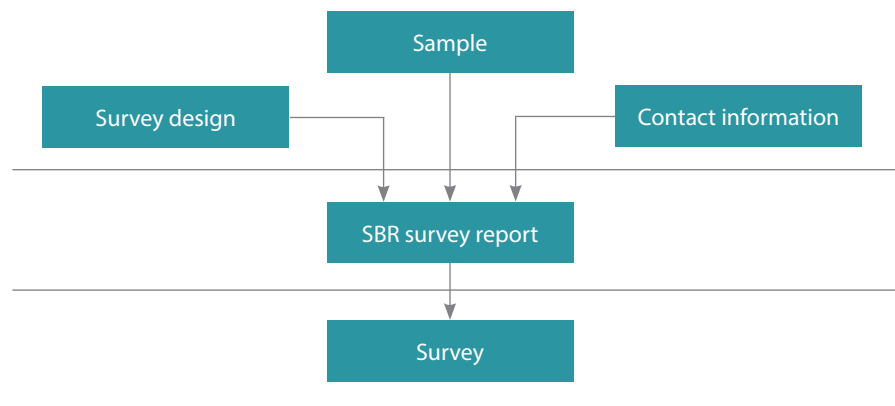
2.6   SBR-based statistics

2.47.   National statistical offices are confronted with increasing demands from governments, international bodies and researchers for business statistics on specific topics such as small areas or sectors of activity, and other aggregates. The information provided must be consistent and comparable at country level, internationally, and for any other geographical area, and also across different economic activities. At the same time, national statistical offices are under pressure to reduce the data-collection costs and response burden.

2.48.   As illustrated in figure 2.6, the SBR can play a key role in providing more information at less cost through the direct use of SBR data. It can be a primary source of information on the structure and demographic characteristics of enterprises. In addition, in combination with other administrative registers or statistical data sources, it can produce additional statistical information for economic analysis. If the SBR is used to integrate a range of data sources, it can also act as a spine for data integration.

Enterprise demography statistics

2.49.   Data on births and deaths of enterprises, their survival rates and the role they play in economic growth and productivity, together with data for tackling socio-

demographic, issues are increasingly requested by both policymakers and analysts. As further discussed in section 9.3, there are requirements for:

- ▸ Data for subpopulations of small and medium-size enterprises.

- ▸ Historical data about enterprises.

- ▸ Data for subpopulations of interest: for example, enterprises with high growth, or with large increases in numbers of employees.

2.50.   In addition, when calculating ratios, conceptual consistency must be maintained between the denominator and numerator populations.

2.51.   The SBR can satisfy all these requirements because it covers the whole population of enterprises. Thus, it can provide regional and small area data, basic characteristics like economic activity code, employment or turnover, and data on births and deaths. In summary, it is an ideal source in terms of coverage and cost for the production and dissemination of enterprise demographic statistics.

Figure 2.6
Statistical business close up statistics
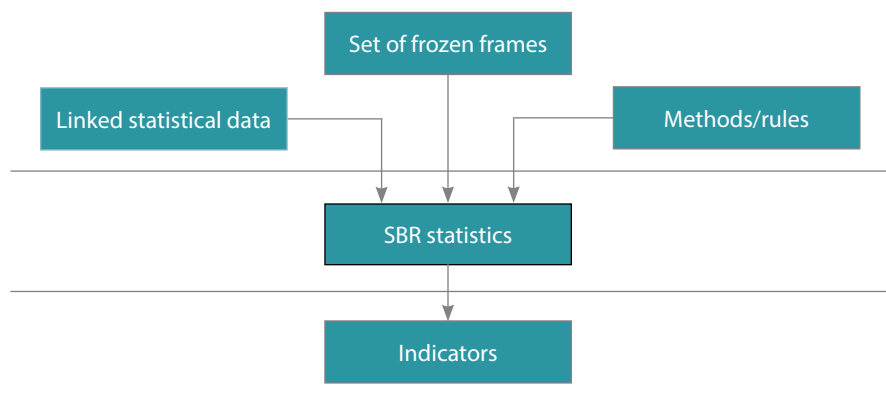
2.52.   The use of the SBR for producing statistics raises similar issues, however, regarding the treatment of data over time as does its use in providing survey frames. The live register is constantly being refreshed. The updates applied over the time period (t) to (t+1) represent not only the actual economic changes that occurred during that period, but also adjustments in the coverage or characteristics of units that resulted from SBR maintenance procedures and that had nothing to do with changes in the real world. For example, based on the results of an SBR improvement survey during the time period from (t) to (t+1) economic activity classifications of enterprises, or their measures of size, or the dates on which they ceased or commenced activity, may be corrected from their former out-of-date, erroneous or missing values. In summary, SBR data cannot be simply regarded as the result of a point-in-time statistical survey, which collects information about the status of an enterprise at time (t) or changes in the enterprise over the period from (t) to (t+1).

2.53.   It may be advisable for statistics derived from SBR data to be based on a *satellite approach*. If parts, or the whole, of one or more frozen frames are extracted from the SBR and linked to data from other sources, and if the resulting information is maintained outside and independent of the SBR environment, then the product is referred

to as a *satellite*. A satellite approach (as further elaborated in section 2.10) is a good way of addressing issues that may arise from corrections of classifications or reference periods as the base population can be well defined and coordinated. There may also be an organizational advantage to a satellite approach in the sense that the task of producing SBR-based statistics from the satellite can be assigned to the organizational unit responsible for economic data production and dissemination rather resting with the SBR unit.

Integration with external trade register

2.54.   External trade statistics do not present any explicit information on the characteristics of enterprises that undertake the trading. The concepts and classifications in external trade statistics differ from those in production statistics. Knowledge of a trading enterprise's profile – economic activity, size, location, and whether or not it belongs to an enterprise group – is important in the analysis of globalization of the national economy and the determinants of this internationalization. The coherent compilation of trade statistics by enterprise characteristics requires the linkage of trade and business registers at the micro level. If this can be achieved, the combination of the key enterprise characteristics and the trade data, such as product code and partner country, offers many opportunities for producing a more complete and diversified view of the structure of both trade and production.[4]

2.55.   Several countries are integrating SBRs with external trade registers. Box 2.4 provides an example of such integration in Tunisia.

[4] For more information on linking the SBR with other data sources on enterprises and also on entrepreneurs, see: United Nations ECE, *Guidelines on the Use of Statistical Business Registers for Business Demography and Entrepreneurship Statistics* (New York and Geneva, 2018), available at https://unece.org/sites/default/files/2021-09/Guidelines%20final.pdf.

Box 2.4
Integration of custom files in the statistical business register in Tunisia

Several sources have been integrated into the SBR in Tunisia to enrich its content, including the customs file. This file contains all the import or export operations per company. It also contains information on the imported or exported product, its value, its weight, and the destination country.

The integration of this file into the SBR is useful for the calculation of detailed statistics on import and export by sector of activity, size of enterprise, legal form, regime or nationality. This integration also makes it possible to estimate and update the content of the SBR, in particular:

> ‣ Estimation of turnover based on the value of exports for only exporting companies.
> ‣ Estimation of the main business activity based on the most exported or imported products.

Prior to 2010, the customs authorities used a company identifier different from the customs code used by the tax authorities. Accordingly, a matching exercise was necessary to integrate the customs file into the SBR based on the company's identification information (name, address, activity, and other elements). From 2010 on, however, the customs authorities began to use the tax code – the same identifier as that used by the tax authorities – as a unique identifier in its files, and the integration of this file in the SBR has now been automated.

2.7   SBR information source

2.56.   The roles discussed in sections 2.5 and 2.6 relate to the production and internal and external use of statistics (in other words, macrodata) based on SBR data. This section concentrates on the provision of SBR micro (individual) data as *open data*, namely, the data that are publicly available for any individuals or organizations to use and that are licensed in a way that makes reuse possible. Access may be made available to:

> ❯   Lists of enterprises – for the use of other government agencies and market research companies so that they can conduct surveys.

> ❯   Ownership and control information, of the "who-owns-whom" variety.

> ❯   Information to support investigations into government programme performance or to support the preparation or monitoring of local, regional or national economic development programmes.

2.57.   As illustrated in figure 2.7, in determining the scope of this role, the national statistical office is dependent on the dissemination infrastructure available. For example, the provision of data through a geographical interface depends on the availability of a suitable geostatistical infrastructure. The national statistical office is also likely to be significantly constrained by the confidentiality and privacy provisions embedded in the laws and regulations governing the dissemination of statistics, as discussed below. A supportive legal framework can open up the possibility of providing external users with access to business data that are already in the public domain, such as: trade name, ownership, economic activity, size, geographical location and contact data, as exemplified by the Mexican DENUE system, which is described in an example below and in chapter 9.

SBR regulatory framework

2.58.   SBR data must be handled in accordance with the regulatory framework that governs the national statistical office. On the one hand, privacy and confidentiality provisions under the statistical legislation have to be observed. On the other hand ,much of the data about enterprises in the SBR is information that the enterprises themselves make public or that is generally considered (within the legal framework in a country) to be of public interest, namely, information to which any citizen should have access.

2.59.   In ideal circumstances a regulatory framework would allow dissemination of the following SBR data:

> ❯   Identification and stratification characteristics: name of enterprise, denomination or corporate name, code and name of the economic activity class, size code based on number of personnel.

> ❯   Geographical location characteristics: street, external and internal number, neighbourhood, postal or zip code, locality.

> ❯   Geographical coordinates of the location: latitude and longitude.

> ❯   Contact characteristics: phone, fax, email, and web page.

> ❯   Date of creation of the unit in the SBR.

2.60.   In practice, the ability to provide this information may require the review and revision of laws, regulations, policies and standards governing both the national statistical office and its administrative sources.

[5] United Nations and others, *System of Environmental-Economic Accounting 2012 – Central Framework* (New York, 2014). Available at http://unstats.un.org/unsd/envaccounting/seeaRev/SEEA_CF_Final_en.pdf.

[6] NACE – Statistical Classification of Economic Activities in the European Community.

Box 2.5
Role of the SBR in the System of Environmental-Economic Accounting (SEEA)

SEEA[5] is the international statistical standard used to measure the environment and its relationship with the economy. The SEEA Central Framework measures the contribution of the environment to the economy and the impact of the economy on the environment. It is a satellite accounting system to the SNA and uses concepts and classifications, for example, for economic activities and products, similar to those employed in the SNA for environmental information. SEEA comprises physical and monetary supply and use tables and asset accounts showing natural resources inputs such as energy and water products and residuals such as air pollution, water emissions, solid waste, oil and gas reserves, environmental investments and subsidies, protection expenditure, and so on.

The SBR supports the compilation of SEEA accounts and tables by providing coordinated populations of statistical units and their characteristics by industry, which can be used to collect information in physical and monetary terms. This allows the derivation of aggregates automatically related to the aggregates of an economic indicator related to the corresponding population of industries used by the national accounts. Accordingly, high quality of the economic activity classification (by ISIC or NACE[6] or both) and the linkage of the SBR with other data sources are of great importance.

Geostatistical information and SBR data consultancy systems

2.61.   In recent years, there has been a growing interest in geospatial analysis. This interest can be attributed to significant increases in the use of geographical information systems (GIS), characterized by the capacity to integrate and use a wide variety of spatial information. Individual address records have become the standard level for spatial investigation in many socioeconomic and planning applications. Because of the increased level of user friendliness and accessibility of GIS packages, geographical coordinates can be assigned to each address in the SBR. The geospatial dimension introduced in the SBR is used as a basis in building a geographical infrastructure data system for geospatial analysis of economic structures.

2.62.   The geospatial information needs to be updated regularly. Ideally, on this basis, the SBR should be able to provide information on the distribution of economic activity and enterprises and establishments by geography. A GIS can present this information as geostatistical cartography. The use of GIS is fundamental in providing users with the necessary data to analyse economic activity in relation to geography.

Figure 2.7
SBR information source
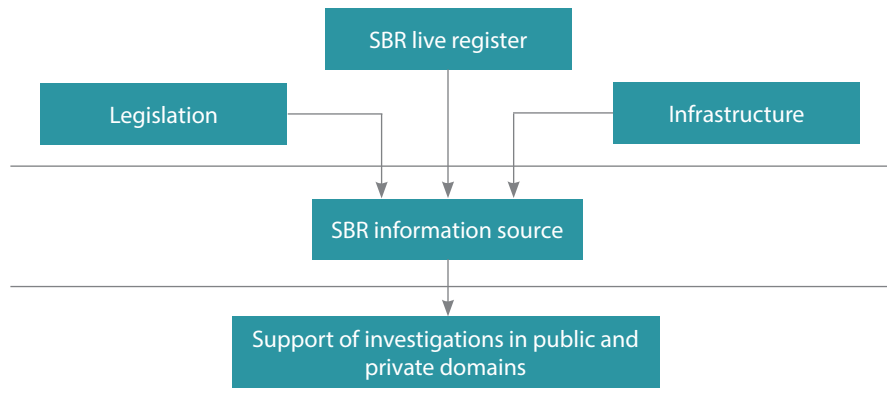
Through a GIS, it is possible to incorporate the many information layers as geostatistical data, including settlements, street axes, block fronts, locality polygons, blocks, external numbers, urban services, reference elements, natural resources, geographical names, hydrographic networks, routes of communication, territory images, relief, limits references, and geodesic references.

Box 2.6
Geostatistical information in Mexico

In Mexico, the National Statistical Directory of Economic Units is a part of the SBR that is available to any user through a free consultation system in INEGI (www.inegi.org.mx/app/mapa/denue/default.aspx). Its geostatistical features have broadened its role as a tool supporting public and private policy developers and decision makers, and also academics and researchers.

2.8   SBR international data exchange

2.63.   In a society that is more and more influenced by globalization, international comparability of economic statistics is becoming increasingly important. Internationally defined indicators based on internationally harmonized populations of statistical units are desirable because the resulting statistics are more comparable when measuring globalization effects. At the same time it is important for producers and users of statistics to understand the potential limitations of the SBR in the context of international comparability.

2.64.   Much effort is spent in the SBR to guarantee coherence, in terms of concepts, methods and quality within a national statistical office and between national statistical offices. In the last few decades, thanks to the activities of supranational and international organizations such as Eurostat, the International Monetary Fund (IMF), OECD, the World Bank, and the United Nations, worldwide coherence in the definition and treatment of enterprises and their statistics, as described in the 2008 SNA and ISIC Rev. 4, has increased considerably. This is important as SBR data are used in studies involving (the structures of) different national economies, as illustrated in figure 2.8.

2.65.   Key factors in improving the quality of international statistics are the exchange and coordination of microdata between countries and the central accessibility of these data, for example, storing them in one central database or at least accessing them through a single gateway. In this context, national SBRs are potential data providers and a network of these could be established. Examples of data exchange between national SBRs include the EuroGroups Register (EGR) and the European System of Interoperable Business Registers (ESBRs), which are described in box 2.7.

2.66.   Data exchange is not only important for international statistics, but also in achieving consistency in statistics dealing with cross-border issues in different countries. For example, a subsidiary of an enterprise in country X should be covered by inward FATS in country Y and by outward FATS in country X.

2.67.   A complex question for a national statistical office is the legal basis for sharing microdata with statistical authorities of other countries or international organizations. International standards regarding the legal basis for microdata sharing are yet to be established. Exchange of data on multinational enterprises can be seriously hampered by the requirements that, first, data are used exclusively for statistical purposes and, second, disclosure of data about individual statistical units is forbidden. The capacity to transfer the data to another country depends upon confidentiality requirements within the sending country. The receiving country is responsible for satisfying whatever conditions are imposed by the sending country and its own confidentiality laws.

Figure 2.8
SBR international data exchange

2.68.   Thus, such exchanges require close cooperation between the partner countries, comparable rules for the use and protection of confidential data, a solid agreement between the countries, and, above all, trust that data are not misused. The forthcoming ECE *Guide to Sharing Economic Data* prepared by the ECE Task Force on Exchange and Sharing of Economic Data analyses specific examples of the sharing of economic data for statistical purposes, identifies obstacles to data sharing and provides guidance, tools and principles to overcome those barriers.[7]

[7]   Task Force on Exchange and Sharing of Economic Data, www.unece.org/statistics/ networks-of-experts/ task-force-on-exchange-and-sharing-of-economic-data. html,

2.9   SBR in the modernization of statistical production and services

2.69.   Development of the SBR is a key to meeting user demands for more, better and more timely statistics. As a unifying tool, the SBR is pivotal in modernizing the statistical production process in line with an integrated statistical architecture. Key elements of current efforts to modernize statistics are efficiency, coherence, interoperability and cooperation. In this context, GSBPM[8] is a standard but flexible tool for describing and defining the common set of business processes that typically constitute statistical production. It comprises three levels: the statistical business process, the eight phases of the statistical business process, and the subprocesses within each phase. The structure of GSBPM is described in chapter 8. GSBPM is being used in developing a common business reference architecture that has the aim of obtaining more coherence in official statistics. The SBR fulfils a backbone role in this respect by integrating several data sources and supporting the collection of economic data. This backbone role can be strengthened when the use of the SBR is seen to support the business processes as described in GSBPM (as further elaborated in chapter 8). Box 2.8 explains the use of GSBPM in the French information system of statistical units (Sirus).

[8]   GSBPM v5.1, https://statswiki.unece.org/display/GSBPM/.

GSBPM subprocess 2.4. Design frame and sample

2.70.   This subprocess only applies to processes which involve data collection based on sampling, such as through statistical surveys. It identifies and specifies the population of interest, defines a sampling frame (and, where necessary, the register from which it is derived), and determines the most appropriate sampling criteria and methodology (which could include complete enumeration). Common sources for a sampling frame are administrative and statistical registers, censuses and information from other sample surveys. This subprocess describes how these sources can be combined if needed. Analysis of whether the frame covers the target population should be performed. A sampling plan should be made.

GSBPM subprocess 4.1. Create frame and select sample

2.71.   This subprocess establishes the frame and specifies the sample. It includes the coordination of samples between repetitions of the same statistical business process, in particular to manage overlap or rotation of the sample. It also includes coordination between samples for different processes by using a common frame to manage overlap and to spread response burden. Quality assurance and approval of the frame and the selected sample are also undertaken in this subprocess, although maintenance of underlying registers is treated as a separate business process.

2.72.   The SBR can be linked to sources other than administrative ones. This can be achieved by storing the identification numbers of units in other sources in the live register, or by integrating special modules in the SBR to be used for matching with other sources. In this case the statistical units in the SBR can be linked to the units in these other sources, which means that their data are available to assist in maintaining the units in the SBR or for use by surveys. To improve interoperability, the SBR should support development of the conditions for exchange of coordinated statistical information among countries, including an international identification service. The SBR may also take advantage of methods and software for SBR maintenance that have been developed by other countries.

Box 2.7
Sirus and GSBPM

Sirus is essentially a harmonized SBR from which the sampling frames for different business surveys are created. The sampling procedure itself lies outside Sirus. Sirus also produces very specific information about the enterprises (the business category) that is incorporated in different dissemination products. Thus, from this point of view, Sirus activities fall within GSBPM subprocesses "4.1 – Collect frame and select sample" and, marginally, "7.2 – Produce dissemination products". Since it is also a crucial instrument for the coherence of the National Institute of Statistics and Economic Studies (INSEE) information system in the field of business statistics, Sirus can also be linked to the overarching quality management process of GSBPM.

More specifically and in terms of their internal functioning, Sirus activities mostly fall into two categories:

› Integration of data from a variety of sources
› Provision of data to other information systems

The data integration processes can be further categorized: their periodicity ranges from daily to annual; the process can be fully automated or require manual intervention; and the information can be accepted without control or manually reviewed and consequently accepted or rejected. Data provision processes can also be distinguished on the basis of their periodicity (weekly to annual, or on demand) and their operation: whether they are automated or not.

In any case, each process has undergone the "Design" and "Build" phases before being implemented. Sirus has also recently been the subject of a quality assessment exercise that relates to the "Evaluate" phase of GSBPM.

Further details of the Sirus process are provided below.

Process conception and construction

Sirus as a whole was carried out as a separate project within INSEE and as such it went through all the subprocesses that constitute the "Specify needs" phase ("1.1 – Identify needs", "1.2 – Consult and confirm needs", "1.3 – Establish output objectives", "1.4 – Identify concepts", "1.5 – Check data availability" and "1.6 – Prepare business case").

For the *data integration processes*, the "Design" and "Build" phases consist of the following:

› In the "Design" phase, the most relevant subprocesses are "2.2 – Design variable descriptions" (specify the file format), "2.5 – Design processing and analysis" (specify the integration procedure) and "2.6 – Design production systems and workflow" (specify the file exchange characteristics). For the sources that require manual verification, the specification of the review procedure can also be classified in subprocess 2.5. There is no need to look more closely here at the subprocess "2.3 – Design collection" because for Sirus the input data are provided by simple file transfers.
› In the "Build" phase, the main subprocesses are "3.2 - Build or enhance process components" (programming of the integration procedures, including manual review), "3.4 - Configure workflow" (setting up the file exchanges), and the test and finalization subprocesses (3.5 to 3.7).

For the *data provision processes*, the "Design" and "Build" phases consist of the following:

› In the "Design" phase, subprocesses "2.1 – Design outputs" (specify the files to be provided) and "2.6 – Design production systems and workflow" (specify the file exchange characteristics) are the most important.
› In the "Build" phase, the central subprocesses are "3.3 – Build or enhance dissemination components" (programming of the extraction procedures), "3.4 – Configure workflow" (setting up the file exchanges), and the test and finalization subprocesses (3.5–3.7).

Figure 2.9 presents a synoptic view of the Sirus process for the phases related to the conception and construction of the system.

Figure 2.9
Sirus process for conception and construction

Quality management / metadata management		
Specify Needs	Design	Build
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components
1.4 Identify concepts	2.4 Design frame & Sample	3.4 Configure workflows
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system
1.5 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process
		3.7 Finalize production system

- Data integration
- Data provision

Process execution

In the process execution (phases 3–8 in GSBPM), it is once again necessary to distinguish between data integration processes and data provision processes.

For *data integration processes*, data collection in Sirus takes the form of file transfers. Consequently, the entire "Collect" phase of GSBPM is in fact limited to a brief "4.3 – Run collection" subprocess (see the last part of paragraph 70 in the GSBPM document). Most of the work after that can be associated with subprocess "5.1 – Integrate data" and in fact most of the upstream activities of Sirus fall under this subprocess.

For the data sources which are accepted automatically in Sirus (Sirene[9], ProfilApp[10], LIFI [11]), the integration process, strictly speaking, stops here. In the case where a manual review of the data set is performed before integration, the process description can be further refined: the reviewing activity would take place in subprocesses "5.3 – Review and validate" and "5.4 – Edit and input".

With the data incorporated in the register through the different data integration processes, Sirus defines new statistical units, such as enterprises from legal units, and new variables, for example, a variable indicating if the unit belongs to the productive sector.

[9] Sirene is the administrative business register (ABR) for all the legal units and their establishments. It is updated daily by the business administrative registrations.

[10] ProfilApp is the register for the enterprises (profiled from enterprise groups).

[11] LIFI is the register for the enterprise groups.

This activity is clearly to be classified under subprocess "5.5 – Derive new variables and units".

Lastly, when the data integration activities are finished, the production of the resulting state of Sirus would fall under subprocess "5.8 – Finalize data files".

The execution of the data provision processes is quite straightforward: the data extractions for the different clients are produced directly from the register (subprocess "6.5 – Finalize outputs"). In some cases, a manual verification is made before the transmission, so a sequence of subprocesses: "6.1 – Prepare draft outputs" – "6.2 - Validate outputs" can be added at the beginning of the data provision process. Figure 2.10 presents a synoptic view of the Sirus execution process.

Figure 2.10
Sirus execution process
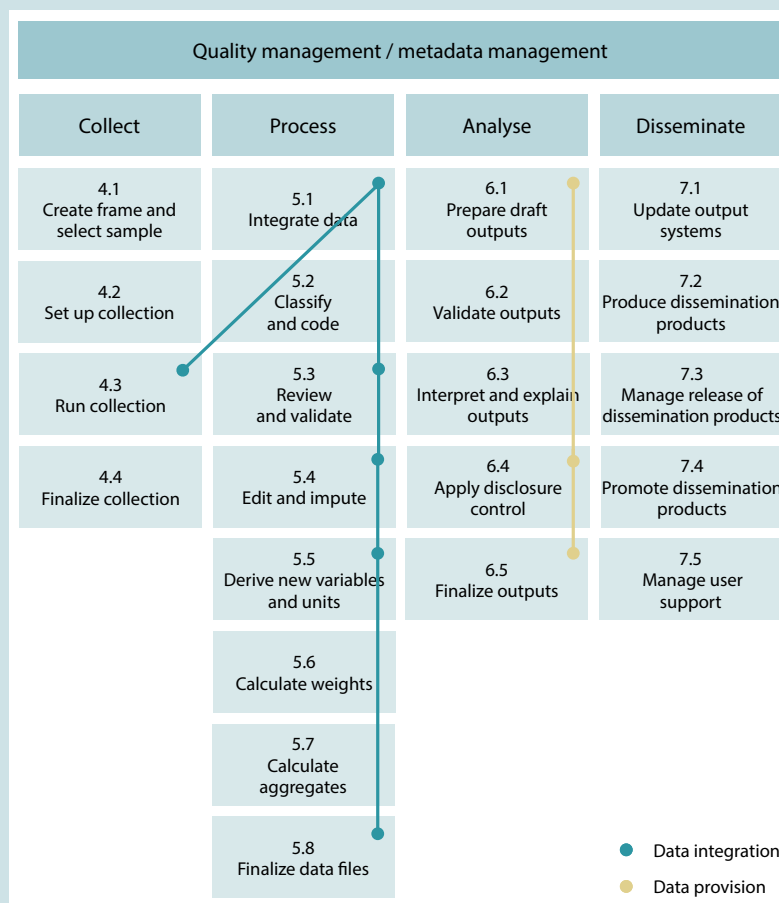
Process evaluation

As part of the implementation of a quality approach, a comparative study of the Sirene and Sirus registers was made in 2015.

Sirene is an inter-administrative register containing legal information that is also included in Sirus, the SBR that complements Sirene. Accordingly, much of this information should be consistent between Sirene and Sirus. This is the first aspect of the task.

There is also a second level of analysis: Sirene determines the legal units in the scope of Sirus in order to ensure that it is comprehensive. In this way it checks that the appropriate

units are sent to Sirus and then actually sends them all. Conversely, it also ensures that Sirene does not send anything to Sirus that should not be sent.

In both of these tasks:

▸ The subprocess "8.1 – Gather evaluation inputs" consists here in selecting specific information from both registers.

▸ The actual work of adjusting information in both registers is part of the subprocess "8.2 – Conduct evaluation".

Once the discrepancies had been identified an action plan was set up to correct them (subprocess "8.3 – Agree an action plan"). The actual task of editing and correcting data must, however, be associated with another phase.

2.73.   Demands for more detailed, timely and coherent data require a modernized national statistical system which produces official statistics more efficiently and in an integrated and internationally coordinated manner. The High-level Group for the Modernization of Statistical Production and Services, established by the Bureau of the Conference of European Statisticians, promotes standards-based modernization of statistical production and services. This calls for a common architecture for data collection, processing and dissemination, in the context of which the SBR will play a key role.[12]

[12] More information about the work of the High-level Group is available on the website of the ECE Statistical Division, www. unece.org/statistics.

Box 2.8
EGR and ESBRs

EGR is the SBR of multinational groups in Europe. EGR was created in cooperation between Eurostat and the national statistical offices of European Union member States and countries of the European Free Trade Association (EFTA), by pooling microdata on legal units, relationships, enterprises and enterprise groups. EGR also acquires data from one commercial data provider in order to ensure coverage for units outside the European Union and EFTA and full consistency of the largest enterprise groups. EGR includes structural economic information, such as economic activity code, turnover and employment, on multinational enterprise groups with an interest in Europe and on their enterprises, and also on their relationships and control structures.

EGR is used for statistics related to globalization at national and European levels in various statistical domains, such as FATS, foreign direct investment (FDI) statistics, short-term statistics, statistics on trade by enterprise characteristics (TEC), national accounts and statistics on small and medium-sized enterprises (SMEs). Statistical users can use EGR as a coordinated frame population, to derive consistent statistical output with improved quality and in measuring global activities of European enterprises part of multinational groups.

More information on EGR may be found at https://ec.europa.eu/eurostat/web/structural-business-statistics/structural-business-statistics/eurogroups-register.

> **Box 2.38** (*continued*)
>
> **ESBRs**
>
> The ESBR project (2013–2020) is streamlining, strengthening and standardizing national SBRs and EGR in the European Statistical System (ESS) with the ultimate aim of making them an efficient interoperable system. It is a continuation of previous EGR and profiling projects carried out in the ESS.
>
> A core goal is the improvement of the EGR statistical frames so that they can provide better quality information on MNEs for globalization statistics. It includes an updated and agreed methodology for European Union profiling that will enable all countries of the ESS to achieve the same view of multinational enterprises; it is based on a collaborative approach with different countries contributing to profiling the same multinational enterprises according to defined roles and responsibilities and supported by a secure platform for sharing confidential data. The ESBRs project includes also practical pilot exercises aimed at testing the proposed solutions with countries of the ESS. ESBRs statistical services are also part of the project's objectives, in the spirit of reusing and sharing tools and solutions.
>
> Additional information on the ESBR project is available at https://ec.europa.eu/eurostat/web/ess/esbr.

2.10   Satellite approach to extend functionality

2.74.   If all the roles described in the previous sections are supported by the SBR, the result is likely to be a complex network of databases and functionalities. One approach that can be used to extend the functionality with minimum complication is to extract parts or the whole of one or more frozen frames from the SBR and afterwards link information from other data sources to its units. The resulting product, which is maintained outside and independent of the SBR, is called an SBR *satellite*.

2.75.   The responsibility for and control of a satellite are separated from the SBR and usually take place in a different environment. While this can adversely affect the coordinating role of the SBR, the big advantage is that much more additional information can be stored and managed without interfering with the basic functions of the SBR. Satellites can, for instance, support international trade statistics, or cooperation with central banks. In such cases, anomalies that might result from integration within the SBR environment can be avoided.

2.76.   The SBR may support the satellite approach in different ways. If the units available in one or more frozen frames are linked to data from sources other than those used in the SBR, then the combined result may be a semi-manufactured statistical product blending statistical and administrative data. The difference from a sample is that, in this kind of satellite, information is linked to all units in the underlying population. This approach is useful, for example, in providing economic indicators on self-employed (income-estimation) or to compile the regional distribution of production factors to support regional accounting information.

2.77.   By linking information collected by different methods, for example by survey and from an administrative source, to the units in one or more frozen frames, a satellite is useful in splitting off and controlling domains where secondary data collection based on administrative sources is not possible or is less efficient. The production of a satellite register is illustrated in figure 2.11, where the population is covered through

three data sources: survey data, data collected from the web and administrative data, all of which are integrated into a single satellite register. The example could also illustrate the traditional approach of merging administrative data for small enterprises and survey data for large enterprises.

Figure 2.11
Production of a satellite register

	Population	Observation
F	P01-RU	Surveyed data
R		
O		
Z	P02-RU	Web-based data
E		
N		
F		
R	P03-RU	Administrative data
A		
M		
E		

P01 = Population 01
P02 = Population 02
P03 = Population 03
RU = Type of reporting unit

Statistical units

Reporting units
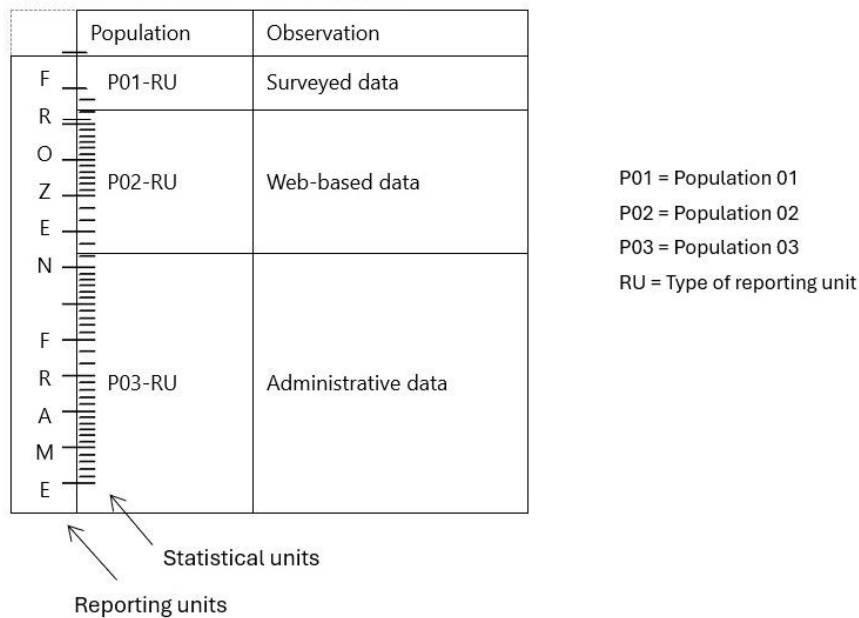
2.78.   It is also possible to analyse and even compensate for over- or under- coverage of units in the live register by comparisons with other sources. Instead of including the other sources in the live register, it may be more convenient (and less complex) to maintain a satellite outside the SBR environment.

2.79.   This kind of approach may be appropriate for units in the governmental sector (S13), or units that belong to the financial sector (S12), or foreign units that belong to the sector representing the rest of the world (S2).

2.80.   Another possible situation in which to set up a satellite is that where different statistical units from those recorded in the live register are used to describe a phenomenon. There may be a case for maintaining descriptions of such units in a satellite to ease the workload on SBR staff and reduce the complexity of the SBR system, even though additional resources may be required to maintain two separate systems. For example, to produce tourism statistics, data are collected on the capacity and the occupancy of accommodation establishments used by tourists, for example, nights spent, arrivals, occupancy rates and number of bedrooms. For this purpose, the data collection at the level of an enterprise unit with a main activity in the tourism groups would be insufficient as many tourist accommodation establishments are provided by enterprises or institutes with another main activity offering accommodation as a secondary activity. For this reason, a special satellite linked to the SBR that records tourism establishments may be a better solution than attempting to extend the functionality of the SBR itself to include such units.

2.81.   In conclusion, it must be reiterated that the main reason for setting up a satellite is to simplify otherwise complicated statistical processes by splitting them off from the SBR, with the aim of simplifying control or spreading responsibility.

Chapter 3
Coverage of the statistical business register

3.1 Introduction

3.1.   This chapter describes the coverage of an SBR and examines the concept from the perspective of the various institutional unit types and sectors of the economy, and it flags key issues that national statistical offices need to consider.

3.2.   In principle, an SBR should record all institutional units in the national economy that are engaged in productive economic activities, namely, activities contributing to GDP. In accordance with 2008 SNA, 1.40, the term "economic production" is understood as referring to "a physical process, carried out under the responsibility, control and management of an institutional unit in which labour and assets are used to transform inputs of goods and services into outputs of other goods and services", and institutional units engaged in production are referred to as enterprises. This chapter examines the three key aspects of SBR coverage:

> *Completeness* is the extent to which the SBR includes all institutional units within the 2008 SNA production boundary.

> *Coverage* is the proportion of total national economic production that the units represent.

> *Content* is the set of characteristics (such as institutional sector, size, location and registration status) of the units included in the SBR.

3.3.   An SBR can be said to provide complete coverage and content if it includes all units producing goods and services, together with all their required characteristics. In reality, complete coverage of all these units within the SNA production boundary is impossible to achieve. For the purposes of international comparisons, however, it is desirable that the coverage of an SBR should meet agreed standards, with completeness as the ultimate aim.

3.4.   During any period, units commence and cease operations. Thus, SBRs should be regularly updated, new units created and units that have ceased economic activities deactivated. Construction and maintenance of the SBR are described in chapters 6 and 7.

3.5.   The SBR provides frozen frames, each of which contains all units in the live register for a given reference period. Frames for economic surveys are derived from frozen frames. Thus the coverage of an SBR for a given reference period for a given survey can be defined as the extent to which the survey frame from the SBR represents the survey target population.

3.6.   This chapter presents considerations that are pertinent in formulating SBR coverage and content decisions and it discusses specific issues for various segments into which the SBR may be divided. It is based on concepts articulated in the 2008 SNA, ISIC Rev. 4, the Eurostat Business Register Recommendations Manual, the Afri-

can Development Bank's Guidelines for Building SBRs in Africa and other sources as footnoted and listed in the References below.

3.2   Key concepts

3.2.1   Production boundaries, residence, institutional sectors and enterprises

3.7.   The concept of "production" is defined in the 2008 SNA, together with descriptions of the production boundaries, which clarify the activities constituting production. The 2008 SNA includes two types of production boundary: the general production boundary and the SNA production boundary.

3.8.   The general production boundary includes production as an activity carried out under the control and responsibility of an institutional unit that uses inputs of labour, capital and goods and services to produce outputs of goods or services. There must be an institutional unit that assumes responsibility for the process of production and owns any resulting goods or knowledge-capturing products, or is entitled to be paid or otherwise compensated for the change-effecting or margin services provided (2008 SNA, para. 6.24).

3.9.   The SNA production boundary is a more restricted version of the general production boundary. It excludes activities undertaken by households that produce services for their own use, except for services provided by owner-occupied dwellings and services produced by employing paid domestic staff (2008 SNA, para. 6.26).

3.10.   The 2008 SNA, in paragraph 4.24, describes the five sectors into which institutional units may be divided. These are non-financial corporations, financial corporations, government units, households and non-profit instituations serving households (NPISHs). Dividing the total economy into sectors enhances the usefulness of the data for the purposes of economic analysis by grouping together institutional units with similar objectives and types of behaviour. Each of the sectors may be divided into subsectors. No single method of subsectoring is optimal for all purposes or all countries. Sectors and subsectors are also used to target or monitor particular groups of institutional units for policy purposes. More details are provided in annex B2.

3.11.   The 2008 SNA refers to institutional units performing the role of economic producers as "enterprises".

3.12.   The institutional sectors listed above include only resident units. The concept of residence in the 2008 SNA is not based on nationality or legal criteria. An institutional unit is said to be a resident unit of a country when it has a centre of predominant economic interest in the economic territory of that country: in other words, when it engages for an extended period (one year or more being taken as a practical guideline) in economic activities on this territory (2008 SNA, para. 2.19). An enterprise is resident in an economic territory when the enterprise is engaged in a significant amount of production of goods or services from a location in the territory (2008 SNA, para. 26.40). The concept of territory is most commonly the area under the effective control of a single government. The economic territory includes the land area, airspace, territorial waters, including jurisdiction over fishing rights and rights to fuels or minerals. The economic territory also includes territorial enclaves in the rest of the world (such as embassies, consulates, military bases and other facilities (2008 SNA, paras. 4.10 and 4.11)).

3.2.2   Market and non-market producers

3.13.   The distinction between market and non-market production is critical from the perspective of the 2008 SNA and hence important for the SBR. Market producers are enterprises that sell all or most of their output at prices that are economically significant (2008 SNA: 4.88). Prices are said to be economically significant if they have a significant effect on the amount that producers are willing to supply and the amounts purchasers wish to buy. Market producers make decisions about what to produce and how much to produce in response to expected levels of demand and expected costs of supply and are exposed to the risks associated with this production. They adjust supply either with the goal of making a profit in the long run or, at a minimum, covering capital and other costs.

3.14.   The core coverage of the SBR comprises units that are market producers and whose principal activity is the production of goods and services. Registered market producers are very important from an economic point of view, and in principle, are not difficult to cover based on available administrative registers.

3.15.   Non-market producers are not likely to respond to changes in economic conditions in the same way as market producers. Their economic behaviour is influenced by the receipt of financial support in the form of transfers such as taxes, grants and donations. Government units and NPIs that provide most of their services at prices that are not economically significant prices are considered non-market producers.

3.16.   The recommended SBR treatment of non-market producers is covered in the discussion on government and NPIs.

3.2.3   Non-observed and observed economies

3.17.   In national accounts the term "non-observed economy" is used to describe activities that, for one reason or another, are not captured in regular statistical enquiries. The reason for this may be that the activity is informal and thus escapes the attention of surveys geared to formal activities; or that the producer is anxious to conceal a legal activity; or that the activity is illegal (2008 SNA, para. 6.39). As noted in the guidance document *Measuring the Non-Observed Economy: A Handbook* (OECD and others, 2002), complete coverage of economic production is a vital aspect of the quality of the national accounts. This is hard to achieve, however, because of the difficulties in recording certain types of productive activities. Activities that are missing from the basic data used to compile the national accounts are said to comprise the non-observed economy. Productive activities may be non-observed because they are informal, household production for own final use, illegal, underground, or due to deficiencies in the basic data-collection system. Making estimates for them in the national accounts is referred to as measurement of the non-observed economy.

3.18.   It is difficult to identify the enterprises that are part of the non-observed economy and thus they are often defined to be outside the planned coverage of the SBR, as further discussed in sections 3.3 and 3.4. In this case their economic activities have to be measured in some way other than by a survey based on the SBR.

3.2.4   Informal sector

3.19.   A national economy can be divided into three parts: the formal sector, the informal sector, and household producing exclusively for own final use. The major portion of the difference between the total economic production of households and what is readily measured by enterprise surveys and administrative sources is the infor-

mal sector, namely, production by enterprises that are considered to be in the informal sector as defined below (para. 3.27). As noted in the 2008 SNA (para. 25.47), the term "sector" in this context "does not have the same basis as the usual use of the word sector throughout the SNA. In the SNA, sectors are made up of complete institutional units; in the context of the informal sector only the productive activities are concerned."

3.20.   The relationship between the non-observed economy and the informal sector is illustrated in figure 3.1, taken from figure 25.1 in the 2008 SNA. The solid circle represents the non-observed economy and the dotted circle the informal sector. The relative size of the three segments shown in the figure varies from country to country.

3.21.   The informal sector represents an important part of the economy in many countries, playing a major role in production, income generation, and employment creation.

Figure 3.1
Non-observed economy and informal sector

NON-OBSERVED   INFORMAL SECTOR

Not observed,
not informal

Observed,
Informal

Not observed,
and Informal
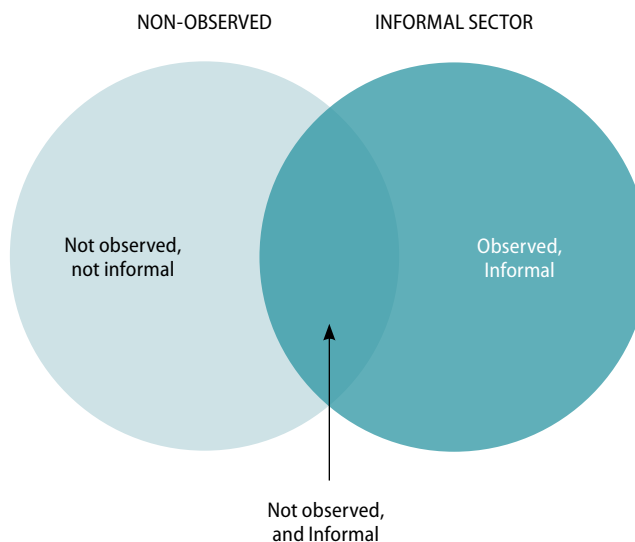
3.22.   The fifteenth session of the International Conference of Labour Statisticians, convened in 1993 by the International Labour Organization (ILO), adopted a resolution concerning statistics of employment in the informal sector. In its paragraph 5 (1), the resolution puts forward a broad definition of the term "informal sector", reflecting the types of production units of which it is composed. The resolution also provides a framework within which each country must formulate its own specific operational definition. Discussion of the informal sector in chapter 25 of the 2008 SNA is consistent with the resolution.

3.23.   The informal sector is broadly characterized as consisting of units engaged in the production of goods or services with the primary objective of generating employment and incomes for the persons involved in the production. These units typically operate at a low level of organization, with little or no division between labour and capital as factors of production and on a small scale. Labour relations – where they exist – are based mostly on casual employment, kinship, or personal and social relations rather than contractual arrangements with formal guarantees (2008 SNA, para. 25.36). The informal sector thus defined excludes households producing exclusively for own final use (2008 SNA, para. 25.41).

3.24. More specifically, the ILO framework requires that the informal sector be defined as a subset of household unincorporated enterprises, comprising those enterprises that produce at least some output for the market and meet one or both of the following conditions:

> ‣ Are less than a specified size in terms of the number of persons engaged, or of employees or of employees employed on a continuous basis

> ‣ Are not registered under specific forms of national legislation, such as factories or commercial acts, tax or social security laws, professional groups' regulatory acts, or similar acts, laws or regulations established by national legislative bodies.

3.25. These criteria provide a framework for a definition of the informal sector, not a single, unambiguous definition. Thus, there are variations across countries in how the informal sector is actually defined. The units included in the informal sector depend upon whether registration or number of employees is used as a criterion, or both, and, if the latter criterion, upon the size boundary for number of employees. It also depends upon whether or not agricultural units are included. In fact, ILO recommends that household unincorporated enterprises that are involved in agricultural activities (ISIC, section A) are identified separately from other economic activities to ensure international comparability and to facilitate the selection and application of appropriate statistical data-collection tools and sample design (2008 SNA, para. 25.46). They are referred to in these Guidelines as "informal agriculture".

3.26. Figure 3.2 shows how the three economic sectors into which the economy may thus be divided relate to the institutional sectors defined in the 2008 SNA, and it indicates which household unincorporated enterprises are considered to be part of the informal sector. This is further elaborated in section 3.3, where a slight distinction between the 2008 SNA and ILO frameworks for the definition of the notion "informal" is described.

3.27. The particular definition of the informal sector recommended in these Guidelines is the set of household enterprises that have market production but are not registered in the administrative source or sources on which the SBR is based. This definition is in accordance with the 2008 SNA and ILO frameworks and provides a clear boundary between the formal and the informal. It means that:

> ‣ The suite of economic surveys based on the SBR and on administrative data measure the formal sector.

> ‣ Household market enterprises not included in the SBR constitute the informal sector (which may itself be divided into informal non-agriculture and informal agriculture).

3.28. Section 3.5.1 contains a discussion on the measurement of the informal sector. More information may be obtained from the document *Measuring Informality: A Statistical Manual on the Informal Sector and Informal Employment*, produced through the collaboration of the United Nations Expert Group on Informal Sector Statistics (Delhi Group) and ILO in 2013. The Manual is a technical guide for national statistical offices and other producers of statistics on the informal sector and informal employment. It presents the ILO international standards along with the guidelines on informal employment adopted by the seventeenth International Conference of Labour Statisticians, in 2003. It also provides practical guidance on implementing the standards, by presenting alternative methods of measurement along with examples based on national experience, and includes guidelines for the dissemination of statistics on the informal sector and informal employment.

Figure 3.2
Identifying units in the informal sector[a]

Institutional sector	Sector of economy		
	Formal	Informal [b]	Household non-market
Non-financial and financial corporations	All	-	-
General government	All	-	-
NPISH	All	-	-
Households	Household market enterprises that are registered or have more than one given number of employees	Household market enterprises that are unregistered or with less than one given number of employees	Household enterprises undertaking production only for own final use

[a] Based on 2008 SNA, figure 25.2.

[b] May be divided into informal non-agriculture and informal agriculture.

3.2.5   Illegal activities

3.29.   The national accounts (2008 SNA, para. 6.43) distinguish two kinds of illegal production:

> ‣ Production of goods or services whose sale, distribution or possession is forbidden by law.

> ‣ Production activities that are usually legal but become illegal when carried out by unauthorized producers, such as, for example, unlicensed medical practitioners.

3.30.   Examples of activities that may be illegal but productive in an economic sense include the manufacture and distribution of narcotics, illegal transport in the form of smuggling of goods and of people, and services such as prostitution. Both kinds of illegal production are included within the production boundary of the SNA provided they are genuine production processes whose outputs consist of goods or services for which there is an effective market demand (2008 SNA, paras. 6.44–6.45).

3.2.6   Underground activities

3.31.   Certain activities may clearly fall within the production boundary of the SNA and be legal, provided certain standards or regulations are complied with, but deliberately concealed from public authorities for such reasons as the following (2008 SNA, para. 6.40):

> ‣ Avoiding payment of income, value added or other taxes.

> ‣ Avoiding payment of social security contributions.

> ‣ Avoiding having to meet certain legal standards such as minimum wages, maximum hours, safety or health standards and others.

> ‣ Avoiding complying with certain administrative procedures, such as completing statistical questionnaires or other administrative forms.

3.32.   Activities of such kinds are referred to as underground activities.[13]

[13] OECD, *Measuring the Non-Observed Economy: A Handbook* (Paris, 2002), section 3.2.2.

3.3   SBR coverage

3.3.1   Type of institutional units

3.33.   The SNA production boundary covers the productive activities undertaken by all institutional units resident in the territory. Thus, national statistical offices should aim to cover as much national production as possible in their SBRs by including all types of institutional units engaged in production. As previously noted, however, complete coverage can never be attained in practice and a more realistic aim is for coverage to meet agreed standards.

3.34.   The following paragraphs set out a detailed discussion on the coverage of the various groups of units in the SBR. Groups of units should not be omitted without an assessment of their economic importance. SBRs should aim to record, at a minimum, all active economic units in the formal economy that are engaged in economic activities contributing to GDP.

3.3.2   Government units

3.35.   The 2008 SNA describes "government units" as unique kinds of legal entities established by political processes that have legislative, judicial or executive authority over other institutional units within a given area (2008 SNA, para. 4.9).

3.36.   Data for general government units are often obtained directly from a single source within the government rather than requiring a direct survey of individual government ministries, departments and agencies.

3.37.   The advantage of including government units in the SBR is that it enables a single system to provide frames for all surveys. Appropriate reporting arrangements can be recorded in the SBR to indicate whether or not the data for a particular survey can be collected from a single centrally located government source rather from the individual government units.

3.38.   The identification of statistical units for government may vary from country to country according to the particular way in which the government is organized. The same general principles for determining the appropriate statistical units can be used for governments, however, as are used in identifying other large complex enterprises. This is further discussed in chapter 7 of these Guidelines.

3.3.3   Corporations

3.39.   In the 2008 SNA (para. 4.7) the term "corporation" covers legally constituted corporations and also cooperatives, limited liability partnerships, notional resident units and quasi-corporations.

3.40.   Units that are market producers and whose principal activity is the production of goods and services constitute the core coverage of the SBR. The subset of these units comprising all registered profit-oriented corporations is the most important one from an economic point of view. The inclusion of these units in the SBR causes, in principle, no coverage difficulty since sources that identify corporations are usually readily available.

3.3.4   NPIs

3.41.   In the 2008 SNA, NPIs are legal or social entities created for the purpose of producing goods and services but whose status does not permit them to be a source

of income, profit or other financial gain for the units that establish, control or finance them (2008 SNA, para. 4.8). Ideally, all NPIs should be included in the SBR. However, some such institutions may not be included in the administrative sources used to maintain the SBR, for example because they are below employment and turnover thresholds.

Box 3.1
Examples of coverage of NPIs: Mongolia and South Africa

Mongolia

The Mongolian SBR, since its establishment in 1998, covers all types of ownership (private entities, government organizations and NPIs). In 2009, the International Classification of Institutional Sectors and the International Classification of Non-Profit Organizations were introduced into the SBR. The introduction of these classifications resulted in the faster and simpler compilation of current and accumulation accounts of the national accounts by institutional sectors. NPIs covered in the SBR are used in the compilation of accounts for NPIs serving households (NPISH).

As the next step in the development of its SBR, the national statistical office of Mongolia is planning to include coverage of the self-employed.

South Africa

Non-profit organizations registered for VAT are contained in the SBR of South Africa. The activity of small non-profit organizations not registered for VAT are not currently included in the SBR, but household surveys could provide relevant information in the future. More information on non -profit organizations in South Africa may be found at www.sars.gov. za/ClientSegments/Businesses/TEO/Pages/default.aspx

When the size of the non-market production is significant in a country, it is critical for non-market units to be properly reflected in the SBR, in order to obtain accurate estimates of changes in consumption and living conditions.

3.3.5   Households

3.42.   Based on the 2008 SNA, households involved in some form of economic production may be referred to as "household (unincorporated) enterprises". As indicated in figure 3.2, they may be divided into four categories:

   (a)   Household enterprises that are defined as belonging to the formal sector, namely, are registered, or have more than a given number of employees

   (b)   Household enterprises without employees belonging to the informal sector (referred to by ILO as "informal own account enterprises")

   (c)   Household enterprises with employees in the informal sector (referred to by ILO as "enterprises of informal employers")

   (d)   Household enterprises undertaking production only for own final use, including owner occupation of dwellings

3.43.   Household enterprises in the categories (b) and (c) may be further subdivided into two subgroups:

   ‣   Market producers selling most or all production, which may themselves be divided into professionals and other

   ‣   Producers for own final use selling some production

3.44.   These subgroups manifest a slight difference in the 2008 SNA and ILO frameworks for defining market production. The 2008 SNA (para. 25.65) defines this difference as follows:

"The ILO treats an enterprise as a market producer if any of the output is sold whereas the SNA requires that most or all of the output be sold. To overcome this difference, it is recommended that a three way split of production be made:

"a. market production according to the SNA criterion whereby most or all output is sold,

"b. output for own final use where some is sold, and

"c. output exclusively for own final use.

"The sum of the first two categories then accords with the ILO guidelines for inclusion in the informal sector as market producers though only the first category is so regarded in SNA terms."

This distinction is illustrated in figure 3.3, which is an expansion of figure 3.2 for the household sector.

3.45.   Ideally, in accordance with the 2008 SNA production boundary, the SBR should include all household enterprises. In practice it is impossible to identify administrative sources that cover all such small units. More realistically, the target coverage for each category of household enterprise is shown in the final row of figure 3.3, which indicates the type of units for which coverage is desirable but for which inclusion in the SBR is conditional on suitable administrative sources being available.

3.46.   The particular informal sector definition recommended in these Guidelines, in paragraph 3.27, is that the informal sector should comprise the set of household enterprises with market production but which are not registered in the administrative source or sources on which the SBR is based. If this definition is adopted then, by extension, the SBR includes only the formal sector enterprises. In terms of figure 3.3, this corresponds to the first and third columns.

3.47.   Enterprises that comprise economic activities of households that are employers of domestic and other personnel should be included in the SBR if they are registered as employers. Their inclusion will enable the SBR to be used in connection with employee or household-based surveys such as the labour-force survey, in which employees of these units are likely to be included.

[14] All rows except the last row are derived from the 2008 SNA, figure 25.4. Enterprises in the government or NPISH sectors are not considered informal.

Figure 3.3
Institutional units in the informal sector and recommended coverage in the SBR[14]

Non-financial and financial corporations			Households		
Those that are registered	Those that are not registered	Those that are registered	Self-employed (Informal own account enterprises)	Unincorporated enterprise with employees (enterprise of informal employers)	Institutional households, households with no unincorporated enterprises, households only undertaking production for own final use (including owner occupation of dwellings)

			Market producers		Producers for own final use		Market producers	Producers for own final use	
			Self-employed professionals	Selling most or all production	Selling some production	Not selling any production	Selling most or all production	Selling some production	Not selling any production
			Informal sector	Informal sector	Informal sector		Informal sector	Informal sector	
Include in SBR	Desirable to include in SBR, but only if suitable administrative sources available	Include in SBR	Desirable to include in SBR, but only if suitable administrative sources available			Do not include in SBR	Desirable to include in SBR, but only if suitable administrative sources available		Do not include in SBR

3.48.   Some household enterprises are lessors, involved only in letting or operating self-owned or leased real estate. If they are not registered as employers they should be excluded from the SBR, first, because they are part of the informal sector, and second, because they occur in large numbers and would inflate the SBR unnecessarily. If an enterprise's letting activities move to a larger scale, the enterprise may well decide to incorporate, in which case the resulting enterprise should be included in the SBR. It is recognized that in some countries the difference between household and incorporated lessors may be difficult to determine on the basis of the administrative sources used for the SBR. In this case, the national statistical office may be compelled to include household lessors (or some of them) in the SBR.

3.3.6   Self-employed professionals

3.49.   Self-employed professionals who are non-employers, such as one person enterprises, are frequently undercovered but tend to play a very important role in several ISIC sections (such as section M – professional, scientific and technical activities – and class 9000). If they are registered with an administrative source they should be included in the SBR. Typical administrative sources are social security files, income tax files, VAT files, files of chambers of commerce and of crafts, government registration files, and lists of doctors. Those that are not registered belong in the self-employed (informal own account enterprises) column in figure 3.3.

3.50.   Some countries may also, or as well, use area enumeration, for example during the course of an economic census, to detect these units, but this approach is not recommended as a source of SBR coverage as it is typically only periodic.

3.3.7   Own-use production

3.51.   The 2008 SNA (paras. 6.28–6.48) specifies that the household production boundary includes the production of goods but excludes the production of most services for own use. The exceptions to the exclusion are services provided by owner-occupied dwellings and services produced by employing paid domestic staff.

3.52.   Household enterprises should not, however, be included in the SBR if the goods and services that they produce are for their own final consumption, such as production from domestic gardens, or invested in the household, such as do-it-yourself activities, as there is no reliable means of maintaining such a list. Thus, the SBR has no role to play in the measurement of household production for own final use. Data about these productive activities are best obtained through surveys of households, not of enterprises.

3.4   Special cases

3.4.1   Agricultural holdings

3.53.   In this context, "agriculture" refers to activities defined in accordance with ISIC section A.

3.54.   It is strongly recommended that agricultural holdings should be included in the SBR provided they are registered. Some countries manage a farm register (possibly including forestry) independently from the SBR. This practice can lead to inconsistency in coverage across economic surveys, as some units may be included in both the SBR and the farm register and others may be excluded from both. It may also result in less effective and efficient updating routines and a greater response burden for the agricultural holdings.

3.55.   The inclusion of agricultural holdings is not only an SBR objective but also an important step towards the integration of agricultural statistics into a comprehensive and coherent system of economic statistics, which is highly desirable, in particular for the purposes of national accounts. Equating a farm with an enterprise is not always straightforward, however, as the farm may be part of a legal entity, or may involve more than one legal entity.

3.4.2   Illegal and underground production

3.56.   The 2008 SNA (para. 3.96) explicitly states that productive illegal activities should be included in national accounts. Illegal actions that fit the characteristics of transactions (notably that have the characteristic of mutual agreement between the parties) are treated in the same manner as legal actions. The production or consumption of certain goods or services, such as narcotics, may be illegal but market transactions in such goods and services have to be recorded in the accounts. If expenditures on illegal goods or services by households were to be ignored on the grounds of principle, household saving would be overestimated and households presumed to obtain assets that they do not in fact acquire. Clearly, the accounts as a whole are liable to be seriously distorted if monetary transactions that in fact take place are excluded. It may be difficult, or even impossible, to obtain data about illegal transactions, but in principle they should be included in the accounts if only to reduce error in other items, including balancing items.

3.57.   Some activities, for example the production and distribution of alcohol, or prostitution, may be illegal in one country but legal in another. Exclusion of illegal production would thus distort international comparisons. Likewise, it would give rise to distortions over time if some activities switched from being illegal to be legal, or the converse.

3.58.   Although, however, there is general agreement on the correctness, in principle, of including illegal activities in the national accounts, in some countries, the difficulties of capturing illegal activities may mean that they are either not well covered

or deliberately ignored on pragmatic grounds (2008 SNA, para. 25.25). The Eurostat *Handbook on the Compilation of Statistics on Illegal Economic Activities in National Accounts snd Balance of Payments*[15] provides guidance on the compilation of illegal activity statistics.

3.59.   In summary, in countries where the value of illegal productive activities is quantitatively insignificant, it is almost certainly a poor use of resources to try to measure them precisely. On the other hand, in some countries the exclusion of illegal activities may lead to significant underestimation of the overall economic activity. Whether or not illegal activities are actually included in the national accounts, an SBR based on administrative registration data cannot be expected to assist in their measurement. If an enterprise undertaking illegal production is registered, it will be included in the SBR but the enterprise will disguise its illegal activities. If the enterprise is unregistered, it will not be included in the SBR. In either case the SBR plays no role.

3.60.   The same applies to underground activities. Unless underground production is very widespread, it is almost certainly a poor use of resources to try to measure it precisely. In any case, the SBR has no role to play as, by definition, the enterprises involved are not registered in the sources on which the SBR is likely to be based.

3.4.3   Extraterritorial organizations and bodies

3.61.   International organizations, embassies and foreign government representations may be divided into two groups:

> ‣   Those whose sites are deemed to form part of the economic territory of another country (such as embassies, consulates and military bases). These units should not be included in the SBR of the host country. They should be included as units in the SBR of the country to which they belong, where they would probably be classified in ISIC section U. For example, the German embassy in Canberra, Australia should be a unit in the German SBR.

> ‣   Those whose sites do not form part of the economic territory of another country. This group includes units such as international organizations (such as the United Nations and its agencies, the European Communities, OECD, IMF, World Bank and others) that are not part of the economic territory of the host country according to the 2008 SNA. For the sake of completeness, however, they should be included in the SBR of the host country, as they do not appear in any other country's SBR. The national statistical office in the host country can decide how to treat them for the purposes of its economic statistics.

3.5   Coverage problems due to the informal sector

3.62.   This section provides guidance on how to handle and measure the informal sector as defined in subsection 3.2.4. and presents some examples of methods followed by countries to cover the informal sector. Care must be taken when expanding the coverage of the SBRs to the informal sector based on the available data sources for SBRs and the updating of the information.

3.5.1   Measurement of the informal sector

3.63.   Given that the majority of people employed in the world work informally (61.2 per cent of global employment),[16] achieving a high quality of coverage in the SBR

[15] Eurostat, *Handbook on the Compilation of Statistics on Illegal Economic Activities in National Accounts and Balance of Payments*, 2018 edition (Luxembourg, European Union, 2018). Available at https://unstats.un.org/unsd/nationalaccount/docs/KS-05-17-202-EN-N.pdf.

[16] ILO, *Women and Men in the Informal Economy: A Statistical Picture.* Third edition (Geneva, International Labour Office, 2018).

is important to ensure that national statistical offices are able to identity and collect statistics on this group to help support policymakers in making informed decisions. An improved and exhaustive SBR that includes units in the informal sector would be useful for enhancing measurements of the informal sector. Work is undergoing on the development of exhaustive business registers by the United Nations Committee of Experts on Business and Trade Statistics to identify, among other things, how the SBR can cover the informal sector.

3.64.   The importance of measuring and understanding informality was highlighted in the 2030 Agenda for Sustainable Development and was also accorded high priority in the ILO Recommendation concerning the transition from the informal to the formal economy, 2015 (No. 204), which specifies that countries should collect data in line with international standards on both the informal sector and informal employment, on a regular basis, in order to monitor and evaluate countries' progress towards formalization.[17]

3.65.   The appropriate method for measuring the informal sector depends upon the question or questions that need to be addressed. If the objective is to monitor the growth of the informal sector employment, then data can be obtained by adding questions to an existing labour force survey. If, however, the objective is to obtain information on the demand by households for goods and services produced by the informal sector, supplementary questions can be included in a household income and expenditure survey.

3.66.   If the objective is to collect detailed structural information on the informal sector, including the number and characteristics of the enterprises involved, their production activities, employment, income generation, capital equipment and relationships with the formal sector, the measurement strategy requires a dedicated informal sector survey in which the non-registered household enterprises themselves and their owners are the observation and reporting units. In this case, establishment surveys and mixed household-enterprise surveys are viable options. The choice depends upon data requirements and resources availability.

3.67.   Establishment surveys are an important source for measuring the informal sector. Traditional establishment surveys, however, use list-based samples constructed on administrative registers, typically a business register. Countries with a large informal sector often do not have a business register. Even where a business register exists, it does not usually cover non-registered household enterprises.

3.68.   In establishment surveys, the use of list-based samples per definition excludes a large part of – or even all – informal household market enterprises. One way to address this problem is to complement the list-based sample with an area-based sample, where all the economic units within the selected areas are identified. The selection of areas (namely, primary sampling units may be made at random from a geographical area-based frame that has been stratified according to, for example, geographical location, concentration of small economic units and their predominant industry. This creates two different complementary frames. The first frame includes the large and medium enterprises that are registered in the administrative source. The second includes the small-scale enterprises that are unregistered and, by definition, informal. The final sample selection can thereby include a census of large enterprises, a sample of medium-sized enterprises selected from the list-based sample and a sample of small sized enterprises selected from the area-based sample, thus covering all enterprises, including informal household market enterprises. This method is more cost-efficient than a full-scale establishment census. It does, however, require a listing of all enterprises within the selected primary sampling units, (including economic activities

[17] ILO, Recommendation concerning the Transition from the Informal to the Formal Economy, 2015 (No. 204), para. 37.

within dwellings and enterprises without fixed location). This approach is relatively costly and complex, due, for example, to sample design requirements, even though it is more limited in scope than a full-scale establishment census.

3.69.   An alternative approach to collecting data on the informal sector is the use of mixed household-enterprise surveys. Such surveys are the most suitable approach when the objective is to collect comprehensive data about the informal sector as a whole and about the various segments of which it is composed. The mixed household-enterprise surveys assume that the informal enterprises are more easily identified through the jobs of the owners of the enterprises than the premises where the activities are taking place. Mixed household-enterprise surveys are based on area sampling and are usually conducted in two stages:

(a)   In the first stage (the household survey component), information from households regarding their activities is collected. Based on this information, owners of informal household market enterprises are identified and form part of the sample used for the next stage. As the information is directly collected from households, all enterprises (with or without fixed location, inside or outside the dwellings) are in scope;

(b)   In the second stage (the enterprise survey component), additional information regarding the enterprise is collected from enterprise owners. This could include information such as the characteristics of the enterprise, expenditure, value of sales, inventories and fixed capital formation.

3.70.   The mixed household-enterprise surveys approach has the advantage that the survey can be conducted either as stand-alone or as modules attached to existing labour-force or other household surveys. For example, connecting the first stage of the mixed household-enterprise survey to an existing survey is a cost-efficient way to build a listing of enterprises in a given area. In addition, it also provides the flexibility that the first and second stages either can be conducted simultaneously – in a single interview – or as two separate interviews. When the information is collected simultaneously, all identified owners of the informal household market enterprises would be included. When divided into two separate interviews, a sample of the identified owners of the informal household market enterprises may be targeted for the establishment survey conducted in the second stage. A mixed household-enterprise survey approach would typically be the most cost-efficient way of collecting data on the informal sector, in particular when it is possible to use existing household surveys.

3.71.   In addition to the direct measurement approaches to measuring the informal sector, indirect estimation methodology can be used to make a rough estimation of the informal sector. Indirect estimation method can be adopted by using different forms of macroeconomic estimates or residual balance approaches based on comparative analysis from different statistical sources.

3.72.   For more information regarding measuring the informal sector, the reader is directed to the above-cited ILO publication *Measuring Informality: A Statistical Manual on the Informal Sector and Informal Employment.* This manual is intended as a technical guide for national statistical offices and other producers of statistics on the informal sector and informal employment.

3.5.2   Country examples

3.73.   The country examples in box 3.2 and figure 3.4 illustrate how these countries deal with the issues of the informal sector.

Box 3.2
Country examples on dealing with the informal sector

Egypt

CAPMAS uses data collected from economic censuses to update its SBR. Egypt conducted its last housing, population and establishment census in 2017. This census was a full enumeration of all the establishments in Egypt, which amounted to some 4.1 million in total. In the economic census, the sample consists of 471,000 establishments at the checking stage (checking the address, the number of employees, the economic activity and the economic sector). After the checking stage, the sample size was reduced to 350,000 establishments following changes in the economic activities and sectors, or the closedown of some establishments.

Census coverage:

▸ Full enumeration of the establishments which have 10 employees or more

▸ 50 per cent of the establishments which have between five and eight employees range

▸ 5 per cent of the establishments which have between one and four employees

▸ Full enumeration of the establishments in each governorate which have so-called "rare economic activities". Rare economic activities are defined as economic activities which have no more than 30 establishments in each governorate.

▸ This sample is then enlarged to represent all establishments in Egypt: some of the establishments which have small number of employees can represent the informal sector, which has the largest proportion in the sample.

▸ The economic census questionnaire has questions about the commercial and industrial registration numbers and the Egyptian industrial union membership. Establishments which do not have this data represent the informal sector, so, by enlarging the sample, CAPMAS can obtain data for the informal sector in Egypt.

CAPMAS carried out an out-of-establishment workers' survey in 2015 (except for agricultural activity), whose sample was based on the sample of the labour force survey. This survey provided data about the economic activities which take place outside the establishments and it is considered to be one of the principal sources for measuring the informal sector. Its data are therefore complementary to the economic census data of the informal sector.

Indonesia

The units in the SBR of Indonesia that are defined to belong to the formal sector include: all non-financial corporations, financial corporations, NPISHs, and government units, and all household unincorporated enterprises that are registered with the administrative sources on which the SBR is based or have been uploaded from the 2016 economic census.

Based on the results of the 2016 economic census, the SBR contains all enterprises from the economic census that have been classified as:

▸ Public enterprise, public corporation, limited corporation

▸ Limited liability partnership

▸ Firm

▸ Cooperative or pension fund

▸ Foundation

▸ Special permit or licence-holder

▸ Subsidiary or representative of a foreign company or organization

Box 3.2 (*continued*)

In Indonesia, the number of small, micro and household businesses is very large. Even based on the 2016 economic census data, more than 90 per cent of the total number of businesses are small, micro and household businesses, in contrast with about 2 per cent for large businesses. For this reason, the scope of the SBR was expanded by adding unincorporated enterprises that have separate financial account. Those were regarded as quasi-corporations. Furthermore, the SBR was expanded with unincorporated medium and large units that were included in the criteria agreed upon for each subject matter areas. For example, for construction statistics, the threshold criteria for the business to be included as unincorporated medium and large units were those units ranked with grade 5, 6, 7 or business scale 3 or 4 in the Indonesian business scale classification system.[18] For large and medium industry statistics, the threshold criteria for the business to be included as unincorporated medium and large units are based on the number of employees and the total output.

After the integration of the 2016 economic census data into the SBR, the next step was to integrate the business data from agricultural statistics. This was necessary because the 2016 economic census only covered businesses and companies outside the agricultural sector.

Before the addition of the economic census data, the SBR data contained the data from the 2006 economic census (large and medium businesses) and local administrative data (administrative data from local administrative sources) gathered by Statistics Indonesia provincial offices, and profiling data. This means that, until now, SBR coverage in Indonesia had come from six data sources, including unincorporated medium and large units and quasi-corporations (see fig. 3.4).

The coverage of the SBR in Indonesia includes not only business units in the formal sector, but also units in the informal sector that have separate financial reports or meet the criteria of being large and medium enterprises as determined by their subject matter areas.

Mexico

In general, economic censuses capture part of the informal economy and their coverage depends heavily on the type of observation unit used in the census. In Mexico, information on the informal sector is obtained through economic censuses and the National Survey for Micro-businesses. The survey is based on the identification of micro-business owners in their households.

State of Palestine

The economy in the State of Palestine is largely composed of establishments that are not registered in the official records. They constitute more than half of the operating establishments. The main reason for the low registration rates is the political situation, which severely limits the data-collection capabilities of the Palestinian Central Bureau of Statistics (PCBS).

The number of operating establishments in the economy was about 150,000 in 2017, based on the findings of the 2017 general census of population and buildings and establishments. The informal sector is a significant component of the economy. As a result, the establishment censuses are carried out every five years. This provides the framework for specialized studies, and also a detailed source of information. Administrative registers are used to identify the large and medium establishments (partly) that entered the market annually and are added to the sampling frame.

The main disadvantages of relying on censuses in the composition of the sampling frame is the cost and the lack of updating at the level of the establishment during the period between the two censuses. There is a gap between the number of establishments in the framework and the number of establishments collected by the census, sometimes as reaching more than 40 per cent, which causes a problem in the estimates of the economic indicators, and these require subsequent revisions.

Box 3.2 (*continued*)

The above problem is addressed by using different sources to measure the appropriate adjusted factor as much as possible to avoid the undercoverage in the data. The sources are:

▸ Labour-force survey (household survey) to estimate the annual increase in the expected number of establishments

▸ Annual establishments survey to update the framework of closed establishments and estimate the rate of change in the change in structural economic activity

The results of the last census (2017) were used to link the available establishment records to different sources. In all, 35 per cent of the units collected could be linked with available records. The main problems arising in this exercise were that there is no single ID used by all sources, different units and recording mechanisms are used, procedures are not updated continuously and the use of the Arabic language for the statistical record linkage methods.

South Africa

Household survey data show that a significant number of persons run non-registered establishments in South Africa. These establishments are relatively small when compared to registered establishments in the formal sector. Informal production tends to be more labour-intensive than formal production. Accordingly, the collection of economic data from small establishments should be investigated, since those small establishments account for a significantly higher fraction of employment in the informal sector than in the formal sector.

The General Household Survey, the Quarterly Labour Force Survey and the Survey of Employers and Self-Employed provide information on household income and expenditure and on informal labour-market activities. These surveys vary to the extent to which information is provided related to the informal sector. Data from administrative sources (when available) can also be used to supplement the survey data on informal sector economic activities.

The main aim of the Survey of Employers and Self-Employed is to provide information about the characteristics of businesses in the informal sector in South Africa and to gain an understanding of their operation and access to services as reported in the survey. The information presented in the Survey of Employers and Self-Employed supplements that in the Quarterly Employment Statistics survey, which collects information from formal sector businesses that are registered for VAT.

The main objectives of the Survey of Employers and Self-Employed are:

▸ To determine the contribution made by businesses which are not registered for VAT to economic growth (the questionnaire has a filter question to ensure that there is no double counting).

▸ To collect reliable data about people running businesses which are not registered for VAT.

▸ To identify the non-income tax-paying and income tax-paying businesses within the non-VAT paying businesses.

▸ To produce comprehensive statistical information about informal sector businesses, at national and provincial levels.

Currently, there is no sampling frame on which to base weights and raising factors for unregistered businesses in South Africa. As a result, the 2013 Survey of Employers and Self-Employed is based on a household survey, consisting of two stages. The first stage involved identifying individuals who are running businesses based on the Quarterly Labour Force Survey. The second stage involved a follow-up, when the owners of these businesses were interviewed, to determine the nature of their businesses.

For each Survey of Employers and Self-Employed survey, the criterion for inclusion depends on whether or not the business is registered for VAT. Only persons who had businesses which were not registered for VAT were included. These businesses are generally excluded from the SBR, which is used by Stats South Africa during surveys to assess the formal economy.

Figure 3.4
SBR coverage and data sources in Indonesia

Sector/activity type	Units	Source recommendation	Current condition in Indonesia	Covered in BPS SBR?
Formal	General government	Administrative data collection	-	Yes
	Financial corporation and quasi-financial corporation	Administrative data collection	2016 economic census	Yes
	Non-financial corporation and quasi non-financial corporation	SBR-based enterprise survey or administrative collection where data available	2016 economic census Subject matter area agricultural enterprise directories	Yes
	NPIs registered as household enterprise		2016 economic census	Yes
Informal	Non-registered non-agricultural household market enterprise	Area-based household enterprise survey, or estimation where value is a small	2016 economic census – unincorporated medium and large units that were included in the agreed criteria based on subject matter areas For example, for the construction statistics subject matter area, the threshold criteria for businesses to be included as unincorporated medium and large units were that they have grade 5, 6 or 7 or business scale 3 or 4	Yes
			Household enterprise from the 2016 economic census	No
Informal agricultural	Non-registered agricultural household market enterprise	Area-based agricultural survey	Subject matter area agricultural enterprise directories	Yes
Own-use production	Non-registered household non-market enterprise	Area-based agricultural survey or estimation where value is small	2013 agricultural census	No
Illegal	Corporation and household enterprise	Ignore, or use experimental estimation within national accounts	-	No
Underground	Corporation and household enterprise	Estimation within national accounts	-	No

3.6 Summary of coverage recommendations

3.74. In principle as many types of enterprises as possible should be included in the SBR. More complete coverage of enterprises means that the SBR can, for example, provide frames for a wider variety of surveys in different sectors of the economy. On the other hand, the more types of enterprises included, the more sources have to be used and the more maintenance has to be undertaken. Thus, decisions by national statistical offices as to which categories of enterprises to include should be based on the appropriate international standards, the availability of suitable sources and the corresponding maintenance costs. A suitable source is one that provides coverage of the units and their basic characteristics, is regularly updated and can be readily acquired and processed by the national statistical office.

3.75. In line with the viewpoint that SBR coverage should be focused on the formal sector, as discussed in section 3.2.4, figure 3.5 summarizes the recommendations for SBR coverage for the various categories of enterprises. It is provided as a guide only. It can be used as a template by which a national statistical office can describe the coverage of its SBR. An example from Australia is presented in figure 3.6.

3.76. Whatever enterprise categories are covered, it is important to ensure that all users, in particular the survey and the national accounts units, are fully aware of the

SBR coverage criteria and the types and overall production of the enterprises that are thus excluded from the SBR. They must also be informed of any planned or unplanned changes in the SBR coverage.

3.77. The survey and national accounts area within the national statistical office should make appropriate provision for measuring or estimating the economic production of enterprises that are not included in the SBR and for informing users of missing coverage. For example, if informal non-agricultural enterprises are excluded from the SBR but their production is thought to be significant, then such production should be measured by means of a two-stage household-enterprise survey. The first stage is a household survey in which households including an enterprise producing for the market are identified. The second stage is a survey of these enterprises. The same approach can be used for informal agricultural enterprises.

Figure 3.5
Recommended SBR coverage

Enterprise type	Recommended SBR coverage	Additional information[a]
Government	Yes	Component of formal sector Inclusion of units in the SBR is preferable but having links in the SBR to data for government units held in an administrative source is acceptable
Financial corporations and quasi-financial corporations (including those controlled by government)	Yes	Component of formal sector
Non-financial corporations and quasi-corporations	Yes	Component of formal sector
NPIs serving households	Yes	Component of formal sector
Registered household enterprises (including agricultural)	Yes	Component of formal sector
Own account,[b] non-registered non-agricultural household market enterprises excluding self-employed professionals	Yes, but only if suitable administrative source available	Component of informal sector If significant but no administrative source, cover through two-stage household-enterprise survey
Own account,[b] non-registered non-agricultural household market enterprises that are self-employed professionals	Yes, but only if suitable administrative source available	Component of informal sector If significant but no administrative source, cover through two-stage household-enterprise survey
Own account,[b] non-registered agricultural household market enterprises	Yes, but only if suitable administrative source available	Component of informal sector If significant but no administrative source, cover through two-stage household-enterprise survey
Non-registered, non-agricultural household market enterprises with employees	Yes, but only if suitable administrative source available	Component of informal sector If significant but no administrative source, cover through two-stage household-enterprise survey
Non-registered, agricultural household market enterprises with employees	Yes, but only if suitable administrative source available	Component of informal sector If significant but no administrative source, cover through two-stage household-enterprise survey
Household non-market enterprises	No	If significant, cover their activities through two-stage household-enterprise survey

[a] Recommended coverage is provided on the assumption that the informal sector is defined as the set of household enterprises that have market production but are not registered in the administrative source(s) on which the SBR is based (see paragraph 3.27).

[b] "Own account" means self-employed with no employees.

Figure 3.6
Example – Australian SBR coverage

Enterprise type	Included in SBR	Additional information
Government	Yes	Government units are eligible for selection in labour collections and other SBR-based collections There is a link between the administrative government finance units and government units on the SBR
Financial corporations and quasi-financial corporations	Yes	These units are eligible for selection in labour collections and other SBR-based collections A link between the surveys of financial information and financial corporation units in the SBR is under development
Non-financial corporations and quasi-corporations	Yes	Micro corporations are separately identified
NPIs serving households	Yes	Micro NPIs are separately identified
Registered household enterprises with greater than a given number of employees (including agricultural)	Yes	Micro units are separately identified
Registered household enterprises with less or equal to than a given number of employees (including agricultural)	No	
Informal own account,[a] non-agricultural household market enterprises excluding self-employed professionals	No	The coverage of the Australian Bureau of Statistics (ABS) SBR is based on entities which have registered for an Australian Business Number. Registration is mandated for businesses above a size cut-off, so this sector is known to be very small
Informal own account,[a] non-agricultural household market enterprises that are self-employed professionals	Yes	
Informal own account,[a] agricultural household market enterprises	No	The coverage of the ABS SBR is based on entities which have registered for an Australian Business Number. Registration is mandated for businesses above a size cut-off, so this sector is known to be very small
Unincorporated enterprises with employees, non-agricultural market producers	Yes	
Unincorporated enterprises with employees, agricultural market producers	Yes	
Household non-market enterprises	No	

[a] Here, "informal" means non-registered and "own account" means self-employed with no employees.

Chapter 4
Units in the statistical business register

4.1 Introduction

4.1.1 Content and references

4.1.   This chapter describes the different types of units from both a conceptual and an operational view. The following broad groups of units are distinguished.

> ▸ *Statistical units* – units defined for statistical purposes, for which information is sought and for which statistics are compiled; there are several types of statistical units.

> ▸ *Legal, administrative and operational units* – legal units recognized by law or society, independently of the persons or institutions that own them; administrative units are units that are defined by a legal unit for the purposes of conforming with an administrative regulation (for example reporting units for accounting purposes or taxation); operational units are defined by a legal unit for the purposes of organizing itself, for example a division, branch, workshop, warehouse, or outlet. All these units are the basis for creating and maintaining statistical units.

> ▸ *Observation and reporting units* - observation units are the units about which data are actually obtained during the course of surveys; reporting units are the units from which data are actually obtained.

4.2.   The set of possible statistical units is usually much broader than those actually incorporated and maintained in the SBR of a country. Which statistical units should be incorporated and maintained in a national SBR depends mainly on the requirements of the business statistics system. Those statistical units which are needed for business statistics should be maintained in the SBR. The SBR should also be the only source for deriving statistical units for survey purposes. Only if this is the case can coherent economic statistics with consistency across sectors, countries and geographical areas, and over time be achieved. While they are not the only requirement for consistent data, statistical units are a basic requirement. For international comparability, harmonized definitions of the statistical units and their harmonized implementation in the national SBRs are required. Reduction in the response burden achieved by collecting data from administrative sources also requires a common understanding of these units.

4.3.   In most cases, a statistical unit is also an observation unit and a reporting unit. The three types of unit may differ, however, in particular when dealing with large and complex enterprises.

4.4.   Links between the different types of units are important. Each national statistical office needs to identify the differences between legal units, administrative units, statistical units and observation units and determine how it is possible to bridge from one to another.

4.5.   The main characteristics of the various types of units and their links are presented in chapter 5. The general structure of the SBR database required to support storage and access of units and their characteristics is presented in chapter 11.

4.6.   The main references for this chapter are the Eurostat Business Register Recommendations Manual, the African Development Bank's Guidelines for Building SBRs in Africa, 2008 SNA and ISIC Rev. 4. Other sources are footnoted and listed in the References below.

4.1.2   Brief presentation of units and problems covered

Statistical units

4.7.   Statistical units about which information is sought and for which statistics are ultimately compiled are the crucial units in the SBR. They should be economically significant, harmonized and have associated with them all the characteristics required for providing frames for economic surveys. They need to be harmonized so as to underpin the production of consistent statistics. They are not always directly observable.

4.8.   The main statistical units used internationally are:

  ‣   Enterprise groups

  ‣   Enterprises

  ‣   Establishments, also called LKAUs

4.9.   Other units with international acceptance are:

  ‣   Kind-of-activity units (KAUs)

  ‣   Local units

4.10.   An MNE is a unit that crosses national boundaries. Until recently individual countries have been interested only in those parts of enterprise groups that operate within the national boundaries. These are referred to as "truncated enterprise groups", while those enterprise groups comprising enterprises that are all contained within the national borders of a country are called "all-resident groups".

Developments in definitions of statistical units in the European Statistical System

4.11.   The definitions of statistical units were laid down in a European Union regulation in 1993. This regulation defines the following statistical units: enterprise, institutional unit, enterprise group, KAU, local unit and LKAU.

4.12.   Over the last few decades, economies have undergone huge structural changes in the organization of production. Both international trade and direct investments in other countries have increased. Enterprises that originally operated only in their home country have become global, and the importance of MNEs has increased considerably. At the same time, the legal complexity of the organization of the enterprises has increased. Enterprises that used to consist of a single legal unit have split their activities across a number of such units. These splits are motivated by tax, liability, tariff and other reasons, not only those directly related to the economic purpose of the enterprise.

4.13.   In order to correctly capture the structure of the economies and their development, it appeared necessary to clarify the statistical units that they used. It was also the case that the statistical units were not uniformly applied across European statistics, resulting in inconsistencies in the data both between member States of the European Union and also between the various statistical domains.

4.14.   It was therefore decided to develop up-to-date and practical rules which should help for a better harmonized interpretation and application of the statistical units. Some of these operational rules will be explored in the present Guidelines, with a view to explaining the underlying concepts.

Legal, administrative and operational units

4.15.   The SBR records data not only about the statistical units but also about all the other types of units that are needed in practice for the creation and maintenance of the statistical units. These types of units include:

- ▸ *Legal unit* – a unit that is recognized by law or society independently of the persons or institutions that own it; this unit is pivotal in linking administrative units and statistical units, because, among other reasons, most administrative registers are based on legal units.
- ▸ *Administrative unit* – a unit specifically designed for the purposes of conforming to an administrative regulation, for example VAT or social security.
- ▸ *Operational unit* – a unit defined by a legal unit for the purpose of organizing itself, for example, a division, branch, workshop, warehouse or outlet.

Observation and reporting units

4.16.   In addition, for the collection of economic data, two more types of units are defined:

- ▸ *Observation unit* – a unit, usually a statistical unit, about which data are obtained during the course of a survey. If a target statistical unit is not directly observable, the corresponding observation unit may be a legal unit or administrative unit linked to the statistical unit. For the most part, observation units are legal or administrative units, but they may also be operational units.
- ▸ *Reporting unit* – a unit from which data about an observation unit are obtained during the course of a survey.

Characteristics of units

4.17.   For each of these types of units, a set of characteristics is held in the SBR, such as economic activity code, size code, institutional sector, whether they are a market or non-market producer, and their region of operations. Such characteristics are of great interest for economic analysis and are also needed as stratification variables in the derivation of survey frames. The main characteristics for each of the types of units are discussed in chapter 5.

Box 4.1
European survey on outward foreign affiliates

In this survey:

- ▸ "Statistical units" are enterprises operating abroad (foreign parts).
- ▸ "Observation units" may be the subsidiaries abroad.
- ▸ "Reporting units" are often the global group heads or the highest level consolidation units, if different.

(All these types of units are described later in the chapter.)

4.2   Types of statistical units

4.18.   As noted above, a statistical unit is a unit about which information is sought and for which statistics are ultimately compiled. It is the unit that provides the basis for statistical aggregates and to which tabulated data refer. Data for a statistical unit may be directly observable, as for example in the case where a statistical unit coincides with a legal unit, or may be derived by splitting or grouping data from observation units using statistical estimation methods, or in some cases may be provided directly by the respondent.[19]

4.19.   In the national SBRs only resident statistical units will be covered. Resident units are units that have their centre of predominant economic interest in the economic territory of that country (see also chap. 3, sect. 3.2). For resident units which, however, are part of a foreign controlled enterprise group some variables and links may be covered in the national SBR. Those additional characteristics could include links with the enterprise group, and control or ownership variables.

4.20.   The various types of statistical units are not independent, but are linked to one another forming a "statistical unit model", that is, a conceptual framework of statistical units and their relationships. These relations are embedded in the SBR by appropriate links. The most important relations are the following:

> ‣   Each enterprise consists of one or more legal units. Very occasionally, in the situation where a split of legal unit is appropriate, an enterprise may contain a part of a legal unit. Thus (except in cases where a legal unit is split), each legal unit is attributed to only one enterprise.

> ‣   In cases where an enterprise consists only of one legal unit, this legal unit corresponds to the enterprise.

> ‣   Each enterprise has one or more local units – or locations. Each local unit is attributed to only one enterprise.

> ‣   In cases where an enterprise has only one local unit, this local unit corresponds to the enterprise.

> ‣   Each enterprise has one or more establishments – LKAUs. Each establishment is attributed to only one enterprise.

> ‣   In cases where an enterprise has only one establishment, this establishment corresponds to the enterprise.

4.21.   The enterprise groups do not cover the entire population of legal units. An enterprise group is defined only in the situation where two or more legal units are linked together by ownership or control. The enterprise groups, therefore, do not cover the legal units which are linked to enterprises that are not part of an enterprise group.

4.22.   When the statistical unit model is to be transformed into an IT data model, each type of statistical unit should have a separate set of identifiers, even if units of one type often correspond to units of another type. In addition, the relations between units should be depicted by separate data elements (links) and not incorporated in the identifiers.

4.23.   In defining standard statistical units, the two most authoritative international sources are the 2008 SNA and ISIC Rev. 4. As described in chapter 3, the 2008 SNA introduces the notions of institutional unit and enterprise, and of statistical units formed by partitioning an enterprise, as follows:

> ‣   An "institutional unit" is defined as an economic unit that is capable, in its own right, of owning assets, incurring liabilities, and engaging in economic activities and in transactions with other entities (2008 SNA, para. 4.2). On the basis of their principal functions, behaviour and objec-

[19] This is often the case for enterprises in a large enterprise group that have been profiled for SBR purposes (as further discussed in sect. 6.6.4) and that may operate solely on the basis of a gentleman's agreement.

tives, institutional units are grouped for national accounts purposes into five mutually exclusive institutional sectors (2008 SNA, para. 4.17), namely non-financial corporations, financial corporations, general government, NPISHs and households.

▸ An "enterprise" may be regarded as the core statistical unit of the 2008 SNA. It is defined as "the view of an institutional unit producing goods and services" (2008 SNA, para. 5.1), as further discussed in section 4.4 below.

4.24.   The 2008 SNA acknowledges that a large enterprise may be engaged in several different types of economic activities, at several different locations. The treatment of such an enterprise for statistical purposes as having a single type of activity at a single location would distort the resulting statistics. Thus, 2008 SNA recommends that large enterprises be partitioned into smaller, more homogeneous producing units (2008 SNA, paras. 5.12–5.14) that collectively represent the enterprise as a whole and can be more precisely classified as follows:

▸ The type of unit resulting from partitioning an enterprise by its different economic activities is a "KAU".

▸ The type of unit resulting from partitioning an enterprise by its different locations is a "local unit".

▸ The type of unit resulting from partitioning an enterprise by both its economic activities and its locations is an "establishment", also called an "LKAU".

Figure 4.1
2008 SNA/ISIC Rev. 4 statistical units model

		Partition by location	
		No	Yes
Partition by activity	No	Enterprise	Local unit
	Yes	Kind-of-activity unit	Establishment (local KAU)

4.25.   This set of statistical units is illustrated in figure 4.1. The enterprise is the basic unit; all other units are defined by partitioning the enterprise.

4.26.   Despite the widespread agreement on the use of 2008 SNA and ISIC Rev. 4, this model has not been universally implemented. In practice, there are deviations from the conceptual framework due to its complexity and the amount of information required for its implementation. Not all types of statistical units are necessarily included in the SBR or used in the various statistical domains.

Recommended basic statistical units model

4.27.   The basic model described in these Guidelines (in sections 4.3–4.5) is in line with that recommended in the SBR Guidelines of the African Development Bank. It is a simplification and extension of the 2008 SNA model and is loosely based on a 2008 European Union business register regulation,[20] which requires that four units – namely, legal unit, enterprise, local unit and enterprise group – need to be maintained in the national SBRs of the States members of the European Union. The model comprises three hierarchically organized statistical units, as illustrated in figure 4.1, namely enterprise group, enterprise and establishment (LKAU).

[20] Business Register (BR) Regulation (EC) No. 177/2008.

Figure 4.2
Recommended statistical units model
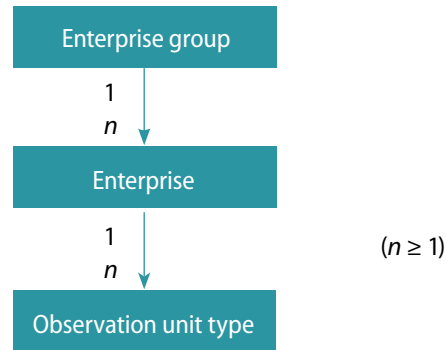
4.28.   The need for an enterprise group is recognized in the 2008 SNA even though it is not defined. The 2008 SNA notes that large groups of corporations, or conglomerates, may be created whereby a parent corporation controls several subsidiaries, some of which may control subsidiaries of their own, and so on. The subsidiaries of a group may be resident in more than one country. For certain purposes, it may be desirable to have information relating to a group of corporations as a whole (2008 SNA, para. 4.51). The notion of an enterprise group is essential in the construction and maintenance of an SBR as it is an important input for delineating and recording the enterprises that belong to a conglomerate, and it is vital for structural statistics as it facilitates linkage of national operating enterprises to the foreign enterprises that control them.

4.29.   The way in which the model fits with the economic and administrative worlds it aims to represent is illustrated in figure 4.3:

‣   The linkage between an administrative unit and a statistical unit is typically through a legal unit but in more complex cases may be through an enterprise group.

‣   Conceptually, an enterprise could also be delineated at the multinational level. In a national SBR, however, the delineation of an enterprise is normally restricted to the national territory.

‣   An establishment is always located in a specific country or a specific region within a country.

‣   Even though an enterprise is usually equal either to a legal unit or to the combination of a number of legal units, the definition allows a legal unit to be split into parts, each part being an enterprise.

4.30.   If required, the basic model can be extended to include local units in accordance with the 2008 SNA model (2008 SNA, para. 5.13).

Figure 4.3
Links between statistical and administrative units
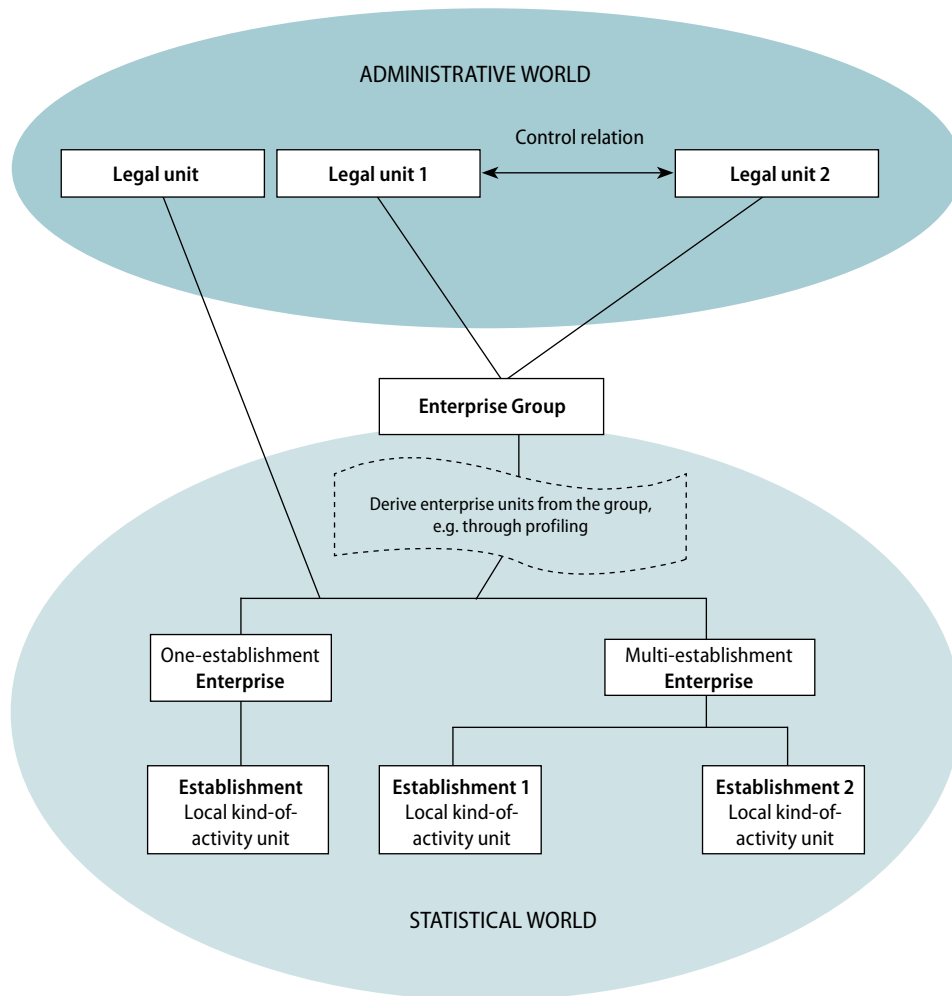
Source: Working Paper on Statistical Business Registers, United Nations Statistical Division and Eurostat (2011).

Box 4.2
European treatment of enterprise groups

The enterprise group is one of the statistical units that need to be implemented and maintained in the SBRs of European Union member States according to the Regulation (EU) 2019/2152 on European Business Statistics.

In paragraphs 21.21–21.26, the Eurostat Business Registers Recommendations Manual states (emphasis added):

> "Whereas the enterprise is an actor in the economy at the level of the production process, with relative autonomy with respect to the allocation and use of its current resources, the **enterprise group** is an actor at a more strategic level taking strategic decisions on behalf of its constituent enterprises (e.g. on product policies, on major expansions, etc.).

> "Since the enterprise group as a whole is subject to demographic events, the statistical unit 'enterprise group' in the register must be maintained according to special continuity rules … for the global, truncated and all-resident groups."

The characteristics of an enterprise group represent properties of the group at global or national (truncated or all-resident) level.

4.3   Enterprise group (and truncated enterprise group)

4.31.   Statistics based on enterprise groups are currently not very well developed, but with increasing globalization, they are progressively becoming of more interest. Statistics using enterprise groups as statistical units are also useful for the analysis of market concentration.

4.32.   Enterprise groups are also important for delineating enterprises within a group. They are needed for compiling data on the impact of multinationals on an individual country's economy, at present mostly in terms of inward and outward FATS and foreign direct investment (FDI) statistics.

4.33.   Ideally, delineation of the enterprises within an enterprise group should be done using a top-down approach, starting from the enterprise group level, in consultation with the global group head or the global decision centre. This is easiest when the global group head or the global decision centre is resident in the country where the delineation is to be done. In cases where only a part of an enterprise group is resident in that country, cooperation and data exchange between the SBRs in all relevant countries are required to delineate all the statistical units belonging to the enterprise group and the corresponding truncated enterprise groups and enterprises in a coherent way. Such cooperation is not easy to achieve, however, and data protection restrictions might make it impossible to exchange microdata between the national statistical office of different countries. This means that the top-down approach may not be feasible and thus a bottom-up approach, starting from legal units belonging to the enterprise group resident in the country, may be the only way to delineate the statistical units.

4.34.   The interest in enterprise groups is not related to the degree of development of an economy but to the shares of the main economic aggregates (such as production, value added and foreign trade) that these enterprise groups control, whether the control is national or foreign. In consequence it is very important to identify enterprise groups, using appropriate methods, even if they are few in number.

4.35.   As previously mentioned, the concept of enterprise group is supplementary to the 2008 SNA model. In fact, consolidation of enterprises into groups is an item on the SNA research agenda. The European Union regulations and research outputs are useful sources to operationalize the concepts related to enterprise group.

4.3.1   Definition of enterprise group (and truncated enterprise group)

4.36.   In the absence of an internationally agreed definition of an enterprise group, reference may be made to the European Union definition and operational rules. Thus, the definition of enterprise group and the associated operational rules below are drawn from documentation agreed upon by the Eurostat Task Force on Statistical Units in the first quarter of 2014.

Definition and explanations

4.37.   An enterprise group is a set of legal units bound together by legal and/or financial links under the same control.

4.38.   Control over a legal unit may be defined as the ability to determine general corporate policy. It can be exercised by, first, owning more than half of the voting shares; second, having the right to appoint or remove a majority of the members of the management; third, having the right to exercise a dominant influence over it; fourth, controlling more than half of the shareholders' voting power of another legal unit directly or indirectly; or, fifth, proving that there is de facto control exercised. Indirect control refers to controlling a legal unit through another legal unit. This also includes also cumulative control, namely, controlling two or more legal units that together own more than half of the voting shares of the legal unit in question.

4.39.   An enterprise group is controlled by its global group head. The global group head is defined as the unit (legal or natural person) which controls all legal units of the group and is not controlled by any other legal unit. The parent legal unit, which is not controlled by any other unit and thus the head of the enterprise group, does not necessarily need to be subject to accounting obligations. This may be the case if the group head is a natural person or the government. Consequently, the unit at the highest consolidation level is not in every case identical with the group head.

4.40.   The subsidiary enterprises of a subsidiary enterprise are considered to be subsidiaries of the parent enterprise.

4.41.   An enterprise group is always controlled by a single global group head. Typically a global group head controls just one enterprise group, but may control more than one. Consolidation of the accounts at a level below the global group head and the existence of several consolidated accounts is an indication that a global group head controls several enterprise groups.

4.42.   The unit carrying out the actual management of the enterprise group is named the "global decision centre". It is not necessarily identical with the global group head.

4.43.   Shares of affiliates' undertakings have to be listed in the balance sheet of a company.[21]

4.44.   All the legal units forming the enterprise group have to be consolidated in its accounts. If the global group head is identical with the highest level consolidating unit, it is also included in the accounts. Legal units not consolidated but controlled by the global group head also form part of the enterprise group.

4.45.   If the global group head does not produce consolidated accounts and reports, the legal unit below the global group head which has to produce consolidated accounts and annual reports is called the "highest level consolidation unit" and it determines the consolidation perimeter of the enterprise group.

[21] Following the European Union Seventh Council Directive and its amendments.

4.3.2   Link between national and global views: truncated enterprise group

4.46.   In practice, an enterprise group may be an "all-resident enterprise group" (otherwise known as a "domestic group" or "national group") or as "MNE" (either domestically controlled or foreign controlled). For the purposes of a national SBR, it is useful to define the concept of a "truncated enterprise group" of an MNE.

Definition

4.47.   A "truncated enterprise group" within a country is the part of an MNE that comprises only the legal units resident in that particular country.

4.48.   If the parent is non-resident it is possible that the truncated enterprise group consists of several seemingly unlinked units and subgroups.

4.49.   A truncated enterprise group may comprise just one unit, a parent or subsidiary, in the national territory. Such truncated enterprise groups may be difficult to identify based on national information only

---

Box 4.3
BPM6 terminology of local and global enterprise groups

The *Balance of Payments and International Investment Position Manual,* sixth edition (BPM6), refers to local and global enterprise groups as follows:

> Groups of enterprises are sometimes identified in defining and classifying direct investment. Although enterprises are the basic unit of economic statistics, a single owner or group of owners may have control of more than one enterprise, so they may act in a concerted way and the transactions between the enterprises may not be driven by the same concerns as "arm's-length" transactions, that is, those with unrelated enterprises.

> Enterprise groups may be either global or local. A global enterprise group refers to an investor and all the enterprises under that investor, whereas the local (or territory-specific) enterprise group refers to an investor and the legal entities under that investor that are resident in the reporting economy. Business accounting may cover groups of related corporate entities (consolidated accounts) including entities that are resident in different economies. However, entities in different economies are not aggregated for macroeconomic statistics that have a focus on an economy. The concepts of global enterprise groups and local enterprise groups are used in the *OECD Benchmark Definition of Foreign Direct Investment.* The global enterprise group is also called a multinational enterprise.[22]

In addition, the *OECD Benchmark Definition of Foreign Direct Investment* provides the following definitions in its Foreign Direct Investment Glossary:

A *multinational enterprise group* consists of all the enterprises located in different economies and under the control or influence of the same owner wherever located.

An *economy-specific enterprise group* consists of all the enterprises located in the same economy and under the control or the influence of the same owner also located in the same economy. Ownership links that are external to the economy are not recognized in the formation of *local enterprise* groups.[23]

More examples of local and global enterprise groups may be found in the *OECD Benchmark Definition,* figure A.4.10 and paragraphs 504 and 505, and in the IMF *Coordinated Direct Investment Survey Guide – 2015,* paragraphs 2.33 and 2.34 and figures 2.1 and 2.2.

---

[22] IMF, *Balance of Payments and International Investment Position Manual*, sixth edition (BPM6) (Washington, D.C., 2009), paras. 4.54–4.55.

[23] OECD, *OECD Benchmark Definition of Foreign Direct Investment*, fourth edition (Paris, 2008), p. 232.

4.3.3   Identifying enterprise groups in practice

Profiling

4.50.   The identification of an enterprise within an enterprise group should, in principle, be made on the basis of the structure and the perimeter of the enterprise group.

In order correctly to delineate the enterprises within an enterprise group it is recommended that, for the most important enterprise groups, the national statistical office contact the enterprise group head or the global decision centre, in order to study the managerial, accounting and legal structure of the enterprise group and the links with legal units. This may lead, in exceptional cases, to splitting a legal unit by allocating its parts to several enterprises within the enterprise group.

4.51.   The process of delineating the statistical units within an enterprise group is called "profiling" (see chap. 6, sect. 6.6.4, for details). It can involve different approaches. In the case of large and complex enterprise groups, "intensive" profiling based on comprehensive contact with the enterprise group is recommended. As the data collected from the large and complex groups usually have great impact on the statistical aggregates, it is useful to devote appropriate resources to these units. For smaller and less complex groups it may be sufficient to delineate the statistical units on the basis of the published accounts and similar information that is directly available.

Group boundary

4.52.   The usual starting point in establishing the boundary (also called "perimeter") around an enterprise group is to consider the consolidated accounts that it contains. In order to delineate enterprise groups based on mutually exclusive accounting groupings, the following rules are applied:

- Consider accounting groups at the highest consolidation level possible (often the global group head).
- Include in an enterprise group the units for which the accounts are entirely integrated in those of the consolidating company, and not the units whose accounts are treated differently (by the equity method in which only the profits and the value of the owned share of equities are consolidated).
- Add majority-controlled units with accounts not included in the overall consolidating by virtue of application of one of the criteria allowed by the Seventh Council Directive,[24] namely, difference in the type of activity or small relative size.
- Discount temporary links of less than a year (which means not including temporarily affiliated units in the enterprise group).

[24] European Union legislation based on article 54 (3) (g) of the Treaty on consolidated accounts.

Special cases of global group heads

4.53.   The global group head of an enterprise group is the enterprise corresponding to the parent legal unit, in other words, the legal unit that is not controlled either directly or indirectly by any other legal unit. The enterprises that it controls are referred to as "affiliates". Affiliate enterprises of an affiliate enterprise are considered to be "subsidiaries" (or "sub-affiliates") of the parent enterprise. There are, however, some forms of cooperative or mutual associations in which the ownership is reversed, the parent enterprise actually being owned by the units of the group.

4.54.   When the global group head does not manage the enterprise group, usually the decisions are taken by the affiliate called the "global decision centre". The SBR has to collect the identification numbers and characteristics of the global group head and of the global decision centre in order to differentiate them clearly when relevant.

4.55.   When the global group head and the global decision centre are not located in the same country, the enterprise group may be counted twice in supranational aggre-

gates, that is, in both the country of the global group head and in that of the global decision centre. Thus, coordination between countries is useful. Such coordination exists in the European Union, with the EGR as one example. The enterprise group is considered to be a resident of the global decision centre's country and not that of the global group head, when the two countries are different.

4.56.   The concept of control implies that natural persons cannot be part of a chain of control unless they are at the top of the structure. Natural persons may, however, not be recorded in the SBR because they are not considered "legal units". The exclusion of natural persons from being group heads can reduce the comparability of enterprise groups.

4.57.   There are a number of other special cases of group heads. The group head may, for instance, be a non-profit body, a trust, or the State, provincial or local government. In general, they are regarded as legal units and can thus be handled according to the normal rules as controlling units, although their involvement in the practical management and decision-making of the enterprise group may vary a great deal. It is also worth noting that the State may control many enterprise groups through one or several ministries and these should not be combined when they operate in different economic activities.

Conglomerates and subgroups

4.58.   An enterprise group may correspond to a conglomerate held together by a network of complex relationships, in which case it frequently covers a very wide range of activities. Subgroups can often be identified within enterprise groups.

4.59.   It is useful to identify all (majority and minority) links between the global group head and a controlled legal unit through the network of subsidiaries and subsubsidiaries. This makes it possible to capture the group's entire organization.

Multinational groups

4.60.   The major challenge is that national statistical offices in general have only very limited information on the non-resident part of MNEs. Statistical surveys and commercial data providers may assist in providing some information. From a national perspective it is difficult, however, to "*see the whole elephant*"[25] or for all the national statistical offices involved to have the same view of an MNE. Nevertheless, this should be the aim of each SBR. Evidently there is more chance of this being possible for an MNE that resides entirely within European Union member States where there are programmes for collaboration than for a group involving countries where no such collaboration exists. The methodology of European profiling addresses this specific challenge (see box 6.11 and annex F1).

4.4   Enterprise

4.61.   The enterprise is the core statistical unit. According to the 2008 SNA and ISIC Rev. 4, it is the level of statistical unit at which all information relating to transactions, including financial and balance-sheet accounts, are maintained, and from which international transactions, an international investment position (when applicable), consolidated financial position and net worth can be derived.[26]

4.62.   The following sections provide a definition of enterprise based on this approach and provide some guidance on how to delineate enterprises.

[25] Richard Barnabé, "Seeing the whole elephant: A proposed experiment on measuring the activities of multinational enterprises", *Statistical Journal of the United Nations Economic Commission for Europe*, vol. 20, No. 2 (January 2003), pp. 173–183.

[26] From ISIC Rev. 4, Introduction.

4.4.1   2008 SNA and ISIC Rev. 4 definitions of enterprise

2008 SNA definition

4.63.   According to the 2008 SNA, an enterprise is the "view of an institutional unit as a producer of goods and services. The term enterprise may refer to a corporation, a quasi-corporation, an NPI or an unincorporated enterprise" (2008 SNA, para. 5.1).

4.64.   In this context, as previously noted, an institutional unit is an "economic entity that is capable, in its own right, of owning assets, incurring liabilities and engaging in economic activities and in transactions with other entities. … There are two main types of units in the real world that may qualify as institutional units, namely persons or groups of persons in the form of households, and legal or social entities" (2008 SNA, paras. 4.2–4.3).

4.65.   The main attributes of institutional units are described as follows:

> "(a)   An institutional unit is entitled to own goods or assets in its own right; it is therefore able to exchange the ownership of goods or assets in transactions with other institutional units;
>
> "(b)   It is able to take economic decisions and engage in economic activities for which it is itself held to be directly responsible and accountable at law;
>
> "(c)   It is able to incur liabilities on its own behalf, to take on other obligations or future commitments and to enter into contracts;
>
> "(d)   Either a complete set of accounts, including a balance sheet of assets and liabilities, exists for the unit, or it would be possible and meaningful, from an economic viewpoint, to compile a complete set of accounts if they were to be required" (2008 SNA, para. 4.2).

ISIC Rev. 4 definition

4.66.   In the Introduction to ISIC Rev. 4, the following definition is provided:

> "An enterprise is an economic transactor with autonomy in respect of financial and investment decision-making, as well as authority and responsibility for allocating resources for the production of goods and services. It may be engaged in one or more productive activities.
>
> "An enterprise may be a corporation (or quasi-corporation), an NPI or an unincorporated enterprise. Corporate enterprises and NPIs are complete institutional units. On the other hand, the term 'unincorporated enterprise' refers to an institutional unit – a household or government unit – only in its capacity as a producer of goods and services" (ISIC Rev. 4, paras. 77–78).

4.67.   The limitations of these definitions are that they do not take into account the relationship between enterprises and the enterprise groups to which they may belong.

4.4.2   European Union definition of enterprise

4.68.   The following definition, from the regulation on statistical units of the European Union,[27] seems more operational, but has not yet been fully applied in the member States:

> "The *enterprise* is the smallest combination of legal units that is an organizational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit" (section III, Definitions).

[27] Council regulation (EEC) No. 696/93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community.

4.69.   Pursuant to this definition, a unit is deemed to be an enterprise if it:

(a)   Operates the necessary factors of production, such as human resources, capital, technology, land and, in particular, management;

(b)   Accesses the necessary controlling system, such as an integrated cost calculation, which covers the main, secondary and ancillary activities of the unit deemed as enterprises;

(c)   Has adequate managerial structures, namely managers that can decide about the production process and about the economic transactions.

4.70.   An enterprise group consisting of two or more legal units may be viewed as a single enterprise if the enterprise group performs its activities under a single management and operates as one organizational unit.

Implementation notes

4.71.   The enterprise is an economic unit that can, in certain circumstances, correspond to a grouping of several legal units. As explained in the regulation, there are two main cases:

"Some legal units, in fact, perform activities exclusively for other legal units and their existence can only be explained by administrative factors (e.g. tax reasons) without them being of any economic significance. … In many cases, the activities of these legal units should be seen as ancillary activities of the parent legal unit they serve, to which they belong and to which they must be attached to form an enterprise."

4.72.   Many of these legal units either have no direct employment, or are the employer of the whole workforce of the group.

4.73.   The enterprise is autonomous as far as production (including allocation of current resources) is concerned. Financial and most structural decisions are taken at a higher level in the enterprise group to which the legal units belong.

4.74.   If a legal unit performs one or more ancillary activities for other legal units within the same enterprise group, it has to be considered to be an ancillary legal unit and, in this case, is not considered an enterprise. The outputs of the ancillary legal unit have to be consolidated as inputs for the other units of the enterprise group and its data have to be consolidated within the enterprises which consume these outputs.

4.75.   In the event that the output of the legal unit, which performs one or more ancillary activities, is only partly consumed by the other legal units, and the ancillary legal unit sells to a third party on a regular basis, it may be treated as an enterprise.

4.76.   As explained in section 4.4.1, an enterprise should have autonomy in respect of financial and investment decisions, in addition to authority and responsibility for allocating resources for the production of goods and services. An enterprise derives its autonomy from the common ownership and control of its resources, irrespective of the number of legal units under which it registers them (ISIC Rev. 4, para. 94). Accordingly, legal units that do not fulfil the autonomy criterion may not be viewed as separate enterprises. They should rather be combined with respective other legal units of the same enterprise group until they together build an enterprise that has autonomy. A legal unit which is not controlled by another legal unit has thus by definition autonomy and should be treated as an enterprise. This will be the case for the great majority of the legal units, as they are not part an enterprise group.

4.77.   Candidates for combining legal units to delineate an enterprise are only those legal units which are part of an enterprise group, which is defined as a set of

legal units bound together by legal and/or financial links under the same control. The delineation of enterprises within an enterprise group could be done by a profiling process. Profiling starts with the legal structure and identifies the operating and organizational group structure; having identified these the criteria for enterprises are applied and this leads to the appropriate enterprises and their constituent legal units. It should be noted that, within enterprise groups, a single legal unit might also be regarded as an enterprise when it fulfils the enterprise criteria. An enterprise group can therefore have one or more than one enterprise.

4.78.   A legal unit or part thereof located in one country may carry out exclusively ancillary activities inside the enterprise group and deliver its services to more than one enterprise of the enterprise group to which it belongs. If the enterprises that receive the ancillary services have locations in one or more other countries, the legal unit providing these services is by convention treated as an enterprise and is classified according to the activity that it is performing.

Box 4.4
Legal units in enterprise groups in France

Some public companies or newly privatized companies (as in rail transport – SNCF – or in postal and financial activities – La Poste) are profiled into several enterprises that have different activities and that may belong to different institutional sectors, for example, financial and non-financial sectors. The global group head, which in these cases is also by far the biggest legal unit of the enterprise group, is organized into different business units (autonomous operational divisions) that are defined as enterprises. Thus, the legal unit has been split in line with the organizational criteria.

4.4.3   Simplified correspondence between legal units and enterprises

4.79.   While it is recommended that all national statistical offices introduce enterprises as distinct from legal units into their SBRs, it is recognized that an enterprise as defined in the 2008 SNA or ISIC Rev. 4, or in the current European definitions, may not be practically implementable in all countries. Data on the basis of which to combine legal units to form an enterprise may not be available, or the SBR system may currently not support the separate recording of legal units and enterprises as distinct units.

4.80.   In the absence of data at enterprise level, it is acceptable that statistical data and demographic data are produced on the basis of legal units. In adopting this approach, however, it is highly recommended that the notion of enterprise be kept separate from that of legal unit, leaving open the future possibility to distinguish and treat individually the most important cases where enterprises that are part of large enterprise groups (whether foreign or domestic) are not coincident with legal units.

Box 4.5
African Development Bank Guidelines on the SBR: recommended definition of enterprise

The African Development Bank Guidelines recommend that legal units and enterprises be defined in a one-to-one correspondence, in the sense that there is one and only one legal unit per enterprise. The benefit of this approach is that it is simple and easy to operationalize. An enterprise is created for each active or potentially active legal unit.

4.5   Establishment

4.5.1   Introductory remarks

4.81.   As discussed in section 4.2 and illustrated in figure 4.1, in partitioning an enterprise by economic activity and geography, the establishment incorporates both the kind-of-activity dimension and the locality dimension. In Europe, the unit equivalent to the establishment is called the "LKAU".

4.82.   For national accounts purposes, industries are defined in terms of establishments. An industry consists of a group of establishments engaged in the same, or similar, types of activity. Although it is possible to use the enterprise as the unit for providing data by industries, some of the resulting industries are likely to be very heterogeneous when enterprises have secondary activities that are very different form their principal activities. It is therefore greatly preferable to partition large, complex enterprises into more homogeneous units, such as establishments, for which production data can be compiled.

4.83.   According to ISIC Rev. 4, the establishment should be used for the analysis of transactions in goods and services and for the compilation of the production account. Furthermore, policymakers interested in investigating regional (state, county, local area) differences want structural data at the subnational level, as provided by data at the establishment (or local unit) level.

4.5.2   ISIC Rev. 4 definition of establishment

Definition

4.84.   ISIC Rev. 4 provides the following definition of "establishment":

> "The establishment is defined as an enterprise or part of an enterprise that is situated in a single location and in which only a single (non-ancillary) productive activity is carried out or in which the principal productive activity accounts for most of the value added" (ISIC Rev. 4, para. 80).

4.85.   The 2008 SNA definition is essentially the same. Although it does not specifically include the "non-ancillary" criterion, this is implied.

4.86.   In the case of most small and medium-sized businesses, the enterprise and the establishment coincide. Large, complex enterprises that are engaged in many economic activities belonging to different ISIC industries are likely to be composed of several establishments provided that production data can be compiled for these smaller, more homogeneous production units.

4.87.   The breakdown of an enterprise into two or more establishments is only required if it has locations situated in two or more different regions and/or it has two or more different, economically significant activities. Thus, only a small number of enterprises will be split into establishments. These, however, will be the larger enterprises and collectively they will have a large share of the resulting industry data. In cases where an enterprise is not split into establishments, the enterprise will be equivalent to the establishment.

4.5.3   Implementation notes

4.88.   The establishment is designed as the statistical unit for which certain data, mainly production, intermediate consumption, investment and employment data, can be surveyed and compiled. Some enterprises, however, do not have bookkeeping practices that allow the delivery of data at the establishment level. In such a case either the enterprise may be considered as consisting of only one establishment (equal to the

enterprise) or the national statistical office may estimate the required economic indicators based on other basic information provided by the enterprise. The first alternative is less desirable as heterogeneous enterprises that are not split into establishments negatively affect the homogeneity of the aggregated industry data.

4.89.   The European regulations do not require the establishment to be directly maintained in the SBR.[28] In addition to the estimation method, data for establishments may also be approximated from local units maintained in, or by means of, administrative registers. Often, but not always, establishments and local units are in a one-to-one relationship, in particular for small and medium-sized businesses.

4.6   Other statistical units

4.90.   The enterprise group, the enterprise and the establishment are the three most important statistical units recommended by these Guidelines. A national statistical office may, however, wish to use either or both of two additional types of statistical unit defined by the 2008 SNA, namely the "KAU" and the "local unit", as previously discussed and illustrated in figure 4.1.

4.91.   The need for the establishment unit is based on the assumption that the aim of the statistical programme is to compile production or production-related statistics that are classified both by activity and by geographical region. In circumstances where precision in the dimensions of both geography and economic activity is not required simultaneously, the KAU or the local unit or both may be more appropriate.

4.6.1   Kind-of-activity unit

2008 SNA definition

4.92.   As defined by the 2008 SNA, a KAU "is an enterprise, or part of an enterprise, that engages in only one kind of productive activity or in which the principal productive activity accounts for most of the value added. Each enterprise must, by definition, consist of one or more KAUs" (2008 SNA, para. 5.12).

Implementation notes

4.93.   The unit is characterized by homogeneity of activity. Compared with the establishment, there is no restriction on the geographical area in which the activity is carried out. When partitioned into two or more KAUs, the resulting units must be more homogeneous with respect to output, cost structure and technology of production that the enterprise as a whole

4.94.   When partitioning an enterprise into more homogeneous KAUs, the only criterion to be considered is the activities of the enterprise. It is irrelevant whether the enterprise has different location or consists of more than one legal unit.

> Box 4.6
> **Example from the European Union statistical units regulation (1993): definition of the LKAU**
>
> In Europe, the unit equivalent to the establishment is called an LKAU. It is the part of a kind-of-activity unit (KAU) that corresponds to a local unit. A local unit is an institutional unit, or part of an institutional unit, producing goods or services situated in a geographically identified place.

[28] See Regulation (EC) No. 177/2008 of the European Parliament and of the Council of 20 February 2008 establishing a common framework for business registers for statistical purposes and repealing Council Regulation (EEC) No. 2186/93.

Box 4.6 (*continued*)

An LKAU groups all the parts of an institutional unit in its capacity as producer contributing to the performance of an activity at the class level (four digits) of the NACE Rev. 2 and corresponds to one or more operational subdivisions of the institutional unit. The institutional unit's information system must be capable of indicating or calculating for each LKAU at least the value of production, intermediate consumption, compensation of employees, operating surplus and employment and gross fixed capital formation.

An LKAU may never belong to two different institutional units.

4.95.   Analogous to the treatment of the establishment, partitioning of the enterprise into KAUs shall primarily be done for large and complex enterprises. The use of estimation methods may also be encouraged in cases where the enterprise is not able (or willing) to provide the economic indicators for the KAUs.

4.6.2   Local unit

2008 SNA definition

4.96.   The 2008 SNA provides the following definition:

> "A local unit is defined as an enterprise, or a part of an enterprise, that engages in productive activity at or from one location. … Location may be interpreted according to the purpose, narrowly, such as a specific address, or more broadly, such as within a province, state, county, etc." (2008 SNA, para. 5.13).

4.97.   The definition refers only to the geographical dimension; it does not refer to the kind of activity that is carried out. The local unit thus defined is conceptually different from a legal or an administrative unit (as described in later sections). It may, however, coincide with such a unit, or be defined on the basis of such units.

4.98.   Each enterprise has at least one local unit, namely the location where the enterprise is registered as a legal unit. Most of the enterprises will have just one location. This is especially the case for smaller enterprises, whereas the larger ones will operate on different locations, thus having more than one local unit. In addition, sole proprietorship enterprises, which are usually quite small, may have more than one location, not only at different addresses in the same community, but also in different communities or regions.

4.99.   More than one activity of an enterprise may take place at a local unit. A local unit may also comprise only ancillary activities. It is recommended that such local units are marked in the SBR. They might also have two classification codes assigned: that of the activity of the local unit and the main activity code of the enterprise. These different codes might be used for different purposes. For business statistics the data will be tabulated on the basis of the main activity of the enterprise and for establishment and regional statistics on the basis of the actual activity performed in the local unit.

Implementation notes

4.100.   For the concept of a local unit, its physical and geographical location is important. A single location is normally best approximated by the corresponding postal address. Several physical locations of the same enterprise within the same community, or within the same region, are to be treated as several local units of that enterprise.

4.101.   The physical allocation of a local unit may be within a building, may correspond to a building or may comprise more than one building. In the last case, the various buildings do not form separate local units if they are physically close together and have a common postal address.

4.102.   A local unit may not be situated in a building at all. In this case other criteria are used in identifying the local unit. A postal address may not exist (for example for a mining site). Identification may still be defined, however, in terms of geographical coordinates or other geographical measures (e.g. the estate number in an administrative land register).

4.103.   Certain economic activities may be performed outside the physical locations of the enterprise, for instance at a customer's address, at fairs, during door-to-door sales, and so forth. These types of location are not to be considered local units. Instead, the sites from which the activities are organized are considered to be the local units. Such activities are treated as if they were carried out at the local unit from which they are organized. The same holds for the activities in transport, where the economic service is transport of goods and persons over the area.

Box 4.7

Example from the European Union statistical units regulation (1993): definition of local unit

The European Union definition of local unit is quite precise about location and employment:

"*The local unit* is an enterprise or part thereof (e.g. a workshop, factory, warehouse, office, mine or depot) situated in a geographically identified place. At or from this place economic activity is carried out for which – save for certain exceptions – one or more persons work (even if only part-time) for one and the same enterprise.

"Explanatory notes

"1. If a person works in more than one place (maintenance or surveillance) or at home, the local unit is taken to be the place from which instructions emanate, from where the work is organized. It must be possible to specify the employment attached to any local unit. However, all legal units that serve as the legal basis for an enterprise or a part thereof must have a local unit which is the registered office, even if nobody works there. Moreover, a local unit can comprise only ancillary activities.

"2. A geographically identified place must be interpreted on a strict basis: two units belonging to the same enterprise at different locations (even within the smallest administrative unit of the Member State) must be regarded as two local units. However, a single local unit may be spread over several adjacent administrative areas, in which case, by convention, the postal address is the determining factor.

"3. The boundaries or the unit are determined by the boundaries of the site, which means for example that a public highway running through does not interrupt the continuity of the boundaries. The definition is similar to the one in paragraph 101 of the introduction to ISIC Rev. 3 in that it concerns localization in the strict sense of the term, but differs from the definition in paragraph 102 in that this strict sense may not vary according to the statistics under consideration. In addition, the criterion of persons working in the unit is normally applied" (sect. III, part F).

4.104.   In the event that the economic activities are performed at the private residence of the entrepreneur, this address is also the address of the local unit of the enterprise.

4.105.   To be defined as a local unit, there should be one or more persons working (even if only part-time) at or from the unit. In the case of seasonal activities, however, a local unit may have unstaffed premises during a certain period of the year.

4.7   Legal unit: pivotal between administrative units and statistics

4.106.   Although enterprises and legal units are usually closely related, the two concepts should not be confused. While legal units are independent in a legal sense, they may not necessarily constitute independent economic units with decision-making autonomy for their activities, in other words, enterprises. Data from legal units without decision-making autonomy may not be comparable with data from legal units with decision making authority.

4.107.   Legal units play a pivotal role in the construction and use of an SBR. Administrative data about economic units are mainly available from legal units. Legal units are often the reporting units from which information is collected about enterprises (as contact and address information is usually available for legal units). Thus the links between the legal units and the corresponding enterprises should be always maintained in the SBR.

4.108.   While legal units are usually the building blocks used in defining enterprises they are not themselves necessarily standardized and comparable across countries since they reflect national administrative and legal requirements, which may differ across countries.

4.109.   While the focus of the SBR is typically businesses, namely, legal units engaged in commercial economic market-oriented production, the SBR may include other legal units, for example NPIs and government units.

4.110.   Administrative regulations usually apply to legal units. In meeting the requirements of a particular regulation, however, and in interacting with the corresponding administrative authority, a legal unit may register and report on the basis of one or more parts of itself, referred to in these Guidelines as administrative units. Administrative data are available for these administrative units.

4.111.   Thus, although in most cases a legal unit has a single administrative unit for a single administrative source, in some cases, often the most important ones, a legal unit has several administrative units for a single administrative source. Furthermore, the sets of administrative units are typically different for different administrative sources. As administrative data are the major input for the SBR, understanding and recording the relationships between the legal units and the various sets of administrative units is vital.

Box 4.8
Typical legal forms in member States of the European Union

‣ *Sole proprietorship*: enterprise owned exclusively by one natural person.

‣ *Partnership*: association of persons who conduct a business under a collective name. It can take the form of a limited partnership.

‣ *Limited liability company*: enterprise that is a joint-stock company, limited partnership with share capital, or private limited company. Harmonized rules at European level governing the publication of accounts for these types of companies are laid down by the Fourth Council Directive.

> Box 4.8 *(continued)*
>
> ‣ *Cooperative society*: body established by law in a country, observing a number of general principles: for example, that it may be entitled to provide services only to members, that profits are often distributed in proportion to members' dealings with the society, and others.
>
> ‣ *Non-profit making body.*
>
> ‣ *Enterprise with other form of legal constitution*: this group includes nationalized industries, publicly-owned enterprises, and state and local authority monopolies.

4.7.1 Definition of legal unit

4.112.   Legal units are recognized by law or by society, independently of the persons or institutions that own them. The characteristics of a legal unit are the following: they own assets, they incur liabilities and they enter into transactions with other entities (contracts). [29]

4.113.   Legal units include:

> ‣ *Legal persons* whose existence is recognized by law independently of the individuals or institutions which may own them, or are members of them
>
> ‣ *Natural persons* who are engaged in an economic activity in their own right as independent self-employed producers

4.114.   A legal unit may be a single entity, controlled or not controlled by another legal unit, or, in some countries, controlled by a group of legal units under common control.

4.115.   Legal units are the core concept in the SBR in the sense that they:

> ‣ Create the operational units by which they manage themselves.
>
> ‣ Register themselves (and possibly administrative sub units) in response to administrative requirements.
>
> ‣ Are modelled in terms of standard statistical units in the SBR.

4.116.   A legal unit always forms, either by itself or sometimes in combination with other (parts of) legal units, the starting point for defining an enterprise.

4.7.2 Market and non-market legal units

4.117.   The legal form of a legal unit is not the only determinant of the institutional sector to which the unit belongs. For example, some legally constituted public sector corporations may be non-market units and hence be part of general government rather than publicly-owned corporations. In addition, some units that are part of general government may be classified as producer units for national accounts purposes and hence represented as enterprises, even though they are not actually independent legal units. They are necessarily recognized as independent legal units. In this case, the corresponding statistical unit in the SBR contributes to the (national, regional) accounts, but is not identified as a legal unit. Thus, both the market- and non-market characteristics of a unit, and its legal form (or quasi-legal form if not recognized as an independent legal unit) are needed to assign the statistical unit to the appropriate institutional sector.

4.118.   Annex B2 indicates the sector classification of enterprises based on legal form and market or non-market orientation.

[29] United Nations Statistical Division and Eurostat. Working Paper on Statistical Business Registers, 2011.

4.7.3   Quasi-corporations

4.119. Some unincorporated enterprises belonging to households or government units may behave in much the same way as corporations. Such enterprises are treated as quasi-corporations, provided they have a complete set of accounts, even though they are not legal units in the strict sense of the term.

4.8   Administrative units

4.120.   Lists of legal units per se are typically not readily suited for statistical purposes. To maintain their SBRs, national statistical offices tend to use data about units from administrative sources.

Definition

4.121.   Administrative units are established by relevant administrative authorities in conjunction with the legal units to which they refer, in order to apply the corresponding regulations. There are many different types of administrative units corresponding to the various administrative regulations. An administrative unit may or may not correspond directly to a legal unit.

Linkage to statistical units

4.122.   The identification numbering system for units in an administrative source often does not correspond to, or is not aligned with, that used in the SBR. In addition, while the legal unit is usually the unit recognized by administrative authorities, legal units may be identified differently in the various administrative databases. Consequently, one of the prerequisites in using administrative data is to establish a rigorous link between the sets of administrative units maintained by administrative authorities and the statistical units in the SBR. Such linkage is necessary to ensure that it is possible to match and combine information from various administrative sources and that there is no duplication in the resulting coverage.

4.123.   As further discussed in section 11.4.10, a unique identifier should be assigned to each statistical unit in the SBR. These identifiers should be linked to the various administrative identifiers. The goal is to ensure that, during the creation and maintenance of statistical units, they are correctly linked to corresponding administrative units.

4.124.   Integration of administrative data into the SBR is greatly facilitated in countries where a single identifier for legal units is used by all administrative authorities in identifying administrative units.

4.9   Units in agriculture and government sectors

4.9.1   Units in agriculture, forestry and fishing

4.125.   Section A of ISIC Rev. 4 covers a range of economic activities: in most countries agriculture is the biggest area; in some countries forestry is a very important industry; and in other countries fishery. ISIC section A also includes activities in support of agriculture, certain agricultural services and post-harvest crop activities, hunting and trapping, logging, and other activities closely related to agriculture, forestry and fishing. The structure and organization of these activities may be quite different from those of businesses ranging from small (family) businesses to large industrial corporations and enterprise groups.

4.126.   Units active in agriculture, forestry and fishing should be covered in the SBR and, if not, there should at least be links to the special farm register if such exists. The statistical system in agriculture may be quite different from the statistical system in the business area owing to the specific nature of agriculture. In particular, the statistical units used in agriculture may differ from those in business statistics. Linking and integrating these two economic areas within the SBR is, however, an important goal for comprehensive and coherent economic statistics.

4.127.   The observation unit in agricultural statistics is mostly an "agricultural holding". This has features similar to those of the statistical units used in business statistics, but may not have a one-to-one relationship directly to any of the units in business statistics. In principle, an agricultural holding may correspond to an enterprise, to a local unit or to an establishment. It could also correspond to a legal unit.

4.128.   The Food and Agriculture Organization of the United Nations defines an agricultural holding as follows:[30]

> "An agricultural holding is an economic unit of agricultural production under single management comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form, or size. Single management may be exercised by an individual or household, jointly by two or more individuals or households, by a clan or tribe, or by a juridical person such as a corporation, cooperative or government agency. The holding's land may consist of one or more parcels, located in one or more separate areas or in one or more territorial or administrative divisions, providing the parcels share the same production means utilized by the holding, such as labour, farm buildings, machinery or draught animals. The requirement of sharing the same production means utilized by the holding, such as labour, farm building, machinery or draught animals should be fulfilled to a degree to justify the consideration of various parcels as components of one economic unit."

4.129.   Whether an agricultural unit in a specific country corresponds more closely to an enterprise, to an establishment, to a local unit, or to a legal unit, it has to be analysed at country level. It has impact on how the corresponding statistical units in the SBR are built and how the SBR is linked to a farm register. The most straightforward case is where an agricultural holding corresponds to an enterprise, as then there is consistency between the level of enterprises in both the agricultural and non-agricultural areas. If a particular enterprise is kept in both the SBR and in the farm register, there is a risk of double counting. To prevent this, checks and clarifications have to be undertaken.

4.130.   In the other cases, where an agricultural holding is not equivalent to an enterprise, statistical enterprises in agriculture might have to be delineated in the SBR in a separate process. If the agricultural holding corresponds to an establishment, the enterprise has to be delineated by linking the establishments that belong to the same enterprise. The relevant establishments might be found not only in the agricultural area, but also in the other economic activities. In this case, it has to be determined whether the resulting enterprise has its principal activity in agriculture or in a non-agricultural industry. An enterprise with its principal activity in agriculture might also produce non-agricultural good and services (for example processed food, transport services, tourist accommodation and others). Conversely, an enterprise with a non-agricultural main activity may also produce agricultural products. In each of these cases, appropriate checks have to be undertaken so that the agricultural part of the SBR is conceptually consistent with the other parts of the SBR and that the units in the SBR do not overlap, and no units are missing.

[30] FAO, *Conducting Agricultural Censuses and Surveys*. FAO Statistical Development Series, No. 6 (Rome, 1996).

4.9.2   Units in government

4.131.   The general government sector consists of institutional units that are non-market producers whose output is intended for individual and collective consumption, and that are financed by compulsory payments made by units belonging to other sectors, together with institutional units principally engaged in the redistribution of national income and wealth. Government units constitute a unique category of legal entities established by political processes that have legislative, judicial or executive authority over other institutional units within a given area.

4.132.   The general government sector consists of two groups of institutional units:

> ‣   Non-market producers whose output is intended for individual and collective consumption and which are financed by compulsory payments made by units belonging to other sectors.

> ‣   Units principally engaged in the redistribution of national income and wealth.

4.133.   The general government sector includes units at all levels of government, from central government to regional and local government, and also social security funds. Public enterprises – which are enterprises owned and controlled by government – are not part of the general government sector. They belong to the financial or non-financial corporate sector. For certain analyses, however, the government sector is lumped together with the public enterprises, and this aggregation is called the "public sector".

4.134.   The units of general government are mainly engaged in section O of ISIC Rev. 4, "Public administration and defence; compulsory social security". Government units are also engaged in other activities, however, mainly in education and health.

4.135.   The organization and structure of the government are quite different across countries, depending on national legislation. Government units may be split into various levels of regional or organizational responsibility. The organizational units may be ministries, departments, offices or bureaux; they may be centrally organized or have regional offices; they may have a separate budget or not; and they may have nearly full, partial or no autonomy in terms of decisions regarding their use of resources.

4.136.   The institutional unit should serve as the common framework for the substantially different government units across countries. As explained earlier, institutional units are units:

> ‣   That are an elementary economic decision-making centre characterized by uniformity of behaviour.

> ‣   That have decision-making autonomy in the exercise of their principal function.

4.137.   An indicator of autonomy is that the unit has a complete set of accounts and responsibility for its budget. Government units do not usually have full autonomy in decision-making, however, even if they have a dedicated budget. Thus, various units may have to be combined to create a unit that closely matches the concept of an institutional unit.

4.138.   For example, a single public school will not have autonomy in the sense of an institutional unit. Autonomy may be achieved only at the level of the ministry of education. In other words, all public schools together with the ministry of education will form a single institutional unit. It may even be the case that the ministry of education does not have full autonomy as, for certain decisions, the ministry of finance also has to agree. Application of this logic in a very strict sense would result in very few governmental institutional units. However, this is quite impractical for SBR purposes as it

would result in a few very big units with a high share of employment but quite differing economic activities.

4.139.   Thus, in defining government units, these Guidelines recommend using a level of aggregation that is a practical balance between the autonomy aspect and the level of detail that is required. It is clear that the government units at the higher organizational levels (such as ministries or departments) will more likely have a higher degree of autonomy than the units organizationally attached to them (such as schools in the example above). In addition, the parliament, high courts and other such bodies are viewed as separate institutional units.

4.140.   In addition to institutional unit, the local unit is also relevant and important for the government sector. Thus, for example, schools under the responsibility of the central government will be recorded as local units of the ministry of education. Of course, if certain schools belong to provincial or local governments these schools will be recorded as local units of either the regional or local government institution responsible for the education sector.

4.10   Branches, ancillary activities, SPEs, franchises, the third or social economy sector, and related concepts

4.141.   This section deals with issues that concern conceptual definitions of units and their practical delineation, that are under current review, scrutiny or research, and that have not been addressed in earlier parts of this chapter. It deals with branches, ancillary units and SPEs, holding companies, head offices and the third or social economy sector. A joint European Central Bank (ECB)/Eurostat/OECD task force has discussed the specific properties of SPEs, holding companies and head offices and proposed harmonized definitions and treatment.

4.10.1   Branch

Definition

4.142.   A branch is a local unit not constituting a separate legal unit in the country where it operates and being dependent on a foreign-controlled enterprise. Branches are treated as quasi-corporate enterprises.

4.143.   Only branches with registered existence (as may be evidenced by a permanent address) are covered by foreign affiliate statistics. In particular, the statistical units for outward FATS are enterprises and all branches abroad that are controlled and managed by an institutional unit resident (in the sense of its global decision centre) in the compiling economy.

Box 4.9
European Union example of branches

A new legal status has been created for the so-called "European companies" or "societates Europaea". This status allows the relevant legal units to operate across the entire European Union without any legal registration in the member States other than the initial one. It enables the place and country of registration to be changed through a simple decision of the board.

> Box 4.9 (*continued*)
>
> In consequence, according to European Union methodology, locations of legal units in, so to speak, "other" countries are branches.
>
> This status, which was rarely used in the recent past, seems now to be developing very quickly. It might create large problems in determining the enterprise group perimeter and country of residence.

4.10.2   Ancillary activities

4.144.   Recommendations regarding the treatment of ancillary activities and the units that perform them are given in the Introduction of ISIC Rev. 4, as follows:

> "59.   Principal and secondary activities cannot be carried out without the support of a number of ancillary activities, such as bookkeeping, transportation, storage, purchasing, sales promotion, cleaning, repair and maintenance, security etc. At least some of these activities are found in every economic entity. Thus, ancillary activities are those that are undertaken to support the main productive activities of an entity by providing goods or services entirely or primarily for the use of that entity.
>
> …
>
> "61.   There are a number of characteristics of ancillary activities that can generally be observed in practice and that help to identify them. The output is always intended for intermediate consumption within the same entity and is therefore usually not recorded separately. Although most ancillary activities produce services, some goods-producing activities may, by exemption, be regarded as ancillary. The goods thus produced, however, may not become a physical part of the output of the main activity (examples are tools, scaffolding ,etc.). Ancillary activities are usually fairly small-scale compared with the principal activity they support.
>
> "62.   If an establishment undertaking ancillary activities is statistically observable, in the sense that separate accounts for the production it undertakes are readily available, or if it is in a geographically different location from the establishments it serves, it may be desirable and useful to consider it as a separate unit and allocate it to the industrial classification corresponding to its principal activity. However, it is recommended that statisticians not make extraordinary efforts to create separate establishments for these activities artificially in the absence of suitable basic data being available.
>
> "63.   Under the definition given in paragraph 59 above, the following activities are not to be considered ancillary:
>
> "(a) Producing goods or services as part of fixed capital formation. The type of units most affected are those doing construction work on the account of their parent unit. This approach is in accordance with the classification in ISIC of own account construction units for which data are available to the construction industry;
>
> "(b) Producing output which, although also used as intermediate consumption by the principal or secondary activity, is for the greater part sold on the market;
>
> "(c) Producing goods that become a physical part of the output of the principal or secondary activity (e.g. the production of boxes, tin cans or the like by a department of an enterprise as packaging for its own products);

"(d) Research and development activities, which are considered part of fixed capital formation in the context of SNA."

4.145.   When a unit is to be classified by economic activity, ancillary activities are not considered. For coding purposes only those activities that form the principal and secondary activities are taken into account. The costs for performing ancillary activities are allocated to the principal and secondary activities of the enterprise.

4.146.   Usually an ancillary activity is not identified as a separate organizational unit or as a separate legal unit. There is, however, an increasing number of legal units that have been created only to perform ancillary activities for the enterprises of an enterprise group.

4.147.   Thus, the general rule is that, since production processes are not usually viable without the support of ancillary activities, these ancillary activities should not be isolated to form distinct entities, even if they are carried out by a distinct legal entity or at a distinct place, and even if separate accounts are kept.

4.10.3   SPEs

4.148.   "Special purpose entity" or "SPE", is the name used in the 2008 SNA for identifying special cases of corporations. In various manuals they are also known as special purpose vehicles (SPVs), international business companies, financing subsidiaries, conduits, holding companies, shell companies, shelf companies, brass plate companies and so on. SPEs are often used as devices to raise capital or to hold assets and liabilities, and usually do not undertake significant production. The 2008 SNA and BPM6 agree that there is no international standard definition of an SPE. Thus, the identification of institutional units that may be described as such units requires consideration of a number of characteristics, including number of employees, physical presence, residency and ownership of financial and non-financial assets.

4.149.   Different ways of treatment are explained in BPM6, which states:

"Although there is no internationally standard definition of SPEs, in economies in which they are important they may be identified separately, according to either a national company law definition, or in terms of a functional definition, possibly referring to their limited physical presence and ownership by non-residents" (para. 4.87).

4.150.   In the 2008 SNA (para. 4.58), a unit defined as an SPE is treated in the same way as any other institutional unit, being allocated to a sector and an industry according to its principal activity, unless it falls into one of the following three categories:

(a)   Captive financial institutions

(b)   Artificial subsidiaries of corporations

(c)   Special purpose units of general government

Box 4.10
ECB/Eurostat/OECD Task Force on Head Offices, Holding Companies and Special Purpose Entities (2013)

The Task Force identified 11 common types of SPEs, namely:

▸ Holding company

▸ Shell company

> **Box 4.10** (*continued*)
>
> ▸ Unit for holding and managing wealth of individuals and families
> ▸ Securitization company
> ▸ Conduit
> ▸ Royalty and licensing company
> ▸ Captive leasing company (including mobile equipment renting company)
> ▸ Factoring and invoicing company
> ▸ Captive insurance company
> ▸ SPE carrying out other financial function
> ▸ Merchanting company
>
> *Source:* Final Report by the Task Force on Head Offices, Holding Companies and Special Purpose Entities (SPEs), ECB, Eurostat OECD, 2013.

4.151. Aware of the difficulty of finding a common international definition, different task forces have tried to define decision trees to be used for allocating units between SPEs and normal units. In particular, the following four main groups have been identified:

   (a)   (Ownership of financial assets (captive financial institutions – 2008 SNA institutional sector S127):

   ▸   Holding companies (ISIC Rev. 4, section K 6420)

   ▸   Trusts, funds and similar financial entities (ISIC Rev. 4, section K 6430)

   ▸   Securitization companies (ISIC Rev. 4, section K 6499 (assuming that the relevant units pass the institutional test, they should be classified as part of institutional sector S125 if they purchase assets on the open markets while raising funds on the open markets. If they do not operate in the open markets on either assets or liabilities, they should be classified in S127.))

   ▸   Captive financial leasing companies (ISIC Rev. 4, section K 6491 (Financial leasing companies operating on open markets are to be classified under institutional sector S125.))

   ▸   Captive insurance and reinsurance companies (ISIC Rev. 4, section K 6512, 6520)

   ▸   Invoicing companies (ISIC Rev. 4, section K 6499).

   (b)   Ownership of non-financial tangible assets:

   ▸   Renting of mobile equipment (ISIC Rev. 4, section N 7730 – institutional sector S11)

   ▸   Merchandising companies (ISIC Rev. 4, section G – 46xx – institutional sector S11)

   (c)   Ownership of non-financial intangible assets:

   ▸   Licensing and royalty companies (ISIC Rev. 4, section M 7490 – institutional sector S11)

   (d)   Others:

   ▸   Offices of airlines in airport hubs abroad (ISIC Rev. 4, section H 5110 – institutional sector S11)

where institutional sector S11 is non-financial corporations, institutional sector S125 is other financial intermediaries except insurance corporations and pension funds, and institutional sector S127 is captive financial institutions and money lenders.

4.152.   As SPEs are currently identified from a national perspective, the following proposals are to be further evaluated:

> ► *Terminology*: it is necessary to clarify the meaning of different labels and to use the same wording for the same things.

> ► *Characteristics*: a clear choice should be made among the terms "no", "few", "little", and "often".

4.153.   The narrowest definition of an SPE could be based on the following features:

> ► No employees

> ► No physical presence

> ► No non-financial assets

> ► Always related to another corporation and their owners are not residents in the territory of incorporation

> ► No other subsidiaries in the country

4.10.4   Distinction between head offices and holding companies

4.154.   Head offices and holding companies are special forms of SPEs. As stated in the above-cited final report by the ECB/Eurostat/OECD Task Force:

> "The statistical analysis of [head offices] and [holiding companies], and their delineation, starts after the [institutional unit] test. Both types of units are often referred to as holding companies, because both of them have relations to other entities, their subsidiaries. However, the relationships are quite different: while a HO exercises managerial control over its subsidiaries, the HC does not undertake any management activities and its principal activity is simply owning a group of subsidiaries."[31]

[31] ECB, OECD and Eurostat, Final Report by the Task Force on Head Offices, Holding Companies and Special Purpose Entities (SPEs), (Paris, 2013), p. 8.

4.155.   From a conceptual point of view the distinction between head offices and holding companies seems clear. However, applying these concepts in practice is complex as head offices and holding companies have common operational characteristics. The following paragraphs provide some guidance.

4.156.   The activities of a head office include the overseeing and managing of other legal units of the enterprise group, thus supporting their day-to-day operations. A holding company is a legal unit that holds the assets of other legal units but does not undertake any management activities. Holding companies are not seen to have autonomy of decision and are combined with one or more producing legal units to form an enterprise.

4.157.   Head offices and holding companies have relations with their subsidiaries. Hence, information on the structure of their balance sheet represents one means of determining whether an entity is a head office or a holding company, or any other type of unit. In order to identify head offices and holding companies, the following practical rule should be applied: a legal unit having at least 50 per cent of its assets in the form of investments in its subsidiaries may be considered to be a head office or a holding company.

4.158.   A head office may have noticeably fewer employees than the legal units that it oversees and manages. That said, however, having zero employment is a clear indication of not being a head office. On the other hand, holding companies that simply hold assets may do so with very few or no employed personnel. Accordingly, employment

thresholds may be applied to differentiate head offices and holding companies. As a general rule, having three or more persons employed is a first indicator that the unit is more likely a head office rather than a holding company.

4.159.   By convention the following rules should help in delineating if a legal unit holding 50 per cent or more of its assets in its subsidiaries should be treated as an enterprise unit or not:

> (a)   A head office or holding company owned by a non-resident parent is to be considered as an enterprise;

> (b)   A head office or holding company owned by multiple owners and not controlled by any other legal unit should be considered as a separate enterprise;

> (c)   For a head office or holding company wholly owned by a single resident unit, having no employees and no compensation of employees are not sufficient criteria for lack of independence. In such case, further investigation is needed;

> (d)   Head offices are always considered to have autonomy of decision.

4.10.5   Franchises

4.160.   Franchise operators, or franchisees, may or may not belong to the same enterprise group. The franchiser is always regarded as a separate enterprise.

4.161.   Franchisees are deemed to be separate from the franchiser because they comprise a complete combination of factors of production, and they run the full entrepreneurial risk. Moreover, although the definition of the enterprise requires autonomy, it allows for this autonomy to be somewhat restricted ("a certain degree of autonomy" is required). Added to which, full accounts tend to be available only at the level of the separate franchisees.

4.10.6   Third or social economy sector

4.162.   A special group of economic institutions belong to the third or social economy sector.[32] The institutions included in this sector include NPIs, cooperatives, mutual societies, and social enterprises:

> ‣   NPIs are defined in section 3.3.4.

> ‣   Cooperatives are organizations formed freely by individuals to pursue the economic interests of their members.

> ‣   Mutual societies, like cooperatives, are organized by individuals seeking to improve their economic situation through collective activity but they differ from cooperatives in that they are mechanisms for sharing risk, either personal or property, through periodic contributions to a common fund.

> ‣   Social enterprises (also known as social ventures) are units that use market means but primarily to serve social purposes, such as employing and training disadvantaged individuals (e.g. persons with disabilities and the long-term unemployed), producing products of special social value or serving disadvantaged persons in other ways.

These units and the compilation of the third or social economy sector are described in the United Nations handbook *Satellite Account on Non-profit and Related Institutions and Volunteer Work*, cited above.

[32] United Nations, *Satellite Account on Non-profit and Related Institutions and Volunteer Work*, ST/ESA/STAT/SER.F/91/Rev.1 (New York, 2018), para. 1.6.

4.163.   The institutions in this sector take diverse organizational and legal forms. For example, NPIs may appear as public charities, public benefit organizations, associations, foundations and non-stock corporations in different countries. Moreover, social enterprises have been legally defined as social cooperatives in Italy, as social impact companies in Luxembourg, as community benefit corporations in the United Kingdom, and as benefit corporations and low-profit limited liability companies (known as "L3C") in the United States.[33] For the purposes of public policy formulation and maximum data utilization, each national statistical office needs to explore the distinct legal forms of all economic institutions as defined under the national legal framework, in order to properly identify these institutions in the SBR, and also to aggregate them for reporting purposes.

4.164.   Despite the diverse organizational or legal forms of NPIs, cooperatives, mutual societies and social enterprises, these economic institutions share common characteristics:

  (a)   They are private (not controlled by government);

  (b)   Their main purpose is to serve a social or public purpose rather than to maximize and distribute returns on invested capital;

  (c)   Engagement in them is voluntary (done by free will and without coercion).[34]

4.165.   Economic institutions in the third or social economy sector, such as NPIs, cooperatives, mutual societies and social enterprises, share common features that set them apart from other units in the corporation sectors with which they are grouped and from similar types of institutions allocated to the general government sector because they are controlled by government. Efforts should be made to distinguish these institutions in the third or social economy sector from other institutions in the SBR.

4.166.   Among the features that distinguish NPIs and other institutions are the following:

  (a)   They have a different objective function from that of standard corporations and are prohibited from distributing profits and income, or significantly limited in their ability to do so, to units that control or finance them. Although they may earn surpluses, the principal purpose of third or social economy sector institutions is not to earn financial returns for investors but to produce goods and services that have some public or social benefit. As a result, they are generally not suitable for investments for profit purposes. Entrepreneurs in such institutions are often driven by social or ideological impulses rather than purely pecuniary motives;

  (b)   Although they often produce goods and services that are sold on the market at economically significant prices, third or social economy sector institutions also typically produce goods and services on a non-market basis that are financed through other means, such as government payments, charitable contributions or volunteer effort. These institutions respond to non-market impulses so as to meet demand for collective goods left unsatisfied by governments and markets, meet new or newly recognized needs, advocate on behalf of citizens and help governments to deliver publicly financed services;

  (c)   Their sources of revenue may differ from those of government or for-profit units. Third or social economy sector institutions may receive a substantial share of their revenue from current transfers that include

[33] Marc J. Lane, *Social Enterprise: Empowering Mission-Driven Entrepreneurs* (Chicago, Illinois: American Bar Association, 2011).

[34] United Nations, *Satellite Account on Non-profit and Related Institutions and Volunteer Work*, ST/ESA/STAT/SER.F/91/Rev.1 (New York, 2018), para. 1.6.

government grants and voluntary contributions that are not available to market producers. This results in different revenue-raising strategies and approaches to the market;

(d) Third or social economy sector institutions are often eligible for tax advantages and exempt from, or eligible for lower rates on, corporate income, sales, property and VATs;

(e) Legal provisions covering the governance, registration, reporting, political participation, commercial activities and other operational or organizational features of third or social economy sector units typically differ markedly from those that apply to for-profit firms. The governing boards of third or social economy institutions are often selected on the basis of one person, one vote by members rather than on the basis of ownership shares, and many countries require such institutions to report publicly on their financial activities annually.

Chapter 5
Characteristics of units

5.1 Introduction

5.1.   This chapter describes the characteristics that should be recorded in the SBR for the various types of units defined and discussed in chapter 4. Collectively, these characteristics identify the units, indicate how they can be contacted and provide economic and geographical information, all time-stamped. They also contain links between different types of units, both within the SBR and across multiple sources of updating information. Attention is paid to the metadata documenting the characteristics and to the time-stamping mechanisms (reference dates and updating dates).

5.2.   The characteristics are grouped and described in accordance with the following classification:

> ▸ Identification and contact characteristics, including not only names, addresses and means of communication (telephone numbers, mail addresses, websites, and the like) but also specific and unique identifiers (even internal identification numbers that enterprises use in monitoring their affiliates and that are not intended for the general public).

> ▸ Demographic characteristics, including unit activity start and end dates, current status and continuity markers.

> ▸ Economic stratification characteristics, such as economic variables and classifications, used for sampling survey frames, grossing up sample data ,and presenting results.

> ▸ Relationships (links) between units of different types within the SBR.

> ▸ Relationships (links) with units in other registers and data sources, including their specific identification numbers.

5.3.   Sections 5.2–5.6 provide general descriptions of the primary characteristics in each of the five groups. Precise details of whether and, if so, how, each characteristic applies to each unit type are provided in annex A.

5.4.   The importance of a characteristic is summarized in terms of its core vs. non-core status:

> ▸ Core characteristics are those that are indispensable in maintaining and/or using the SBR. Most characteristics described in this chapter are core characteristics.

> ▸ Non-core characteristics are those that are optional and can be added in a later stage, such as geographical coordinates to supplement addresses.

5.5.   Identification numbers are very important characteristics because they are essential in merging microdata from various sources, including from administrative files.

5.6.   Some characteristics need more explanation and guidance regarding use than others. These include economic activity, sector and regional classifications, and the

size classes. Not only must these basic classification concepts be defined, but the processes by which their values are acquired also need to be described.

5.7.   Chapters 6 and 7 detail the data sources used to establish and maintain the values of the characteristics. For every value, the associated metadata (including source, date of update and reference date) should be recorded, as further discussed in section 10.6. The primary reference source for chapter 5 is the Eurostat Business Registers Recommendations Manual.

5.2   Identification and contact characteristics

Type of unit

5.8.   *Purpose:* To identify the type of a unit.

5.9.   *Definition:* Chapter 4 lists and defines the main types of units that may be included in the SBR. These are "enterprise group", "enterprise", "establishment", "KAU", "local unit", "legal unit" and "administrative unit".

5.10.   As noted in chapter 4, some enterprise groups operate and are managed on a global scale. For purely national purposes they should be truncated to the national territory. Identification of truncated enterprise groups does not involve anything more than splitting the multinational enterprise groups on a national basis.

Type of enterprise group

5.11.   *Purpose:* To identify the type of enterprise group.

5.12.   *Definition:* The type options are "all-resident" (all enterprises of the group are resident in the same country), "multinational domestically controlled" (its global decision-centre is in the country compiling the SBR), and "multinational foreign controlled" (its global decision-centre is outside the country compiling the SBR).

Role of enterprise within enterprise group

5.13.   *Purpose:* To identify the role of an enterprise within an enterprise group.

5.14.   *Definitions:* The type options are "management/control unit", "global group head" (controlling unit), "global decision centre" (managing unit), "highest level consolidation unit", and "other".

Identification number(s)

5.15.   *Purpose:* To identify the unit and to enable it to be linked to other units in the SBR and in administrative and other statistical sources.

5.16.   It is preferable that no information on the content of the unit is embedded in the identification number, as further discussed in section 11.4.10.

Name(s)

5.17.   For legal persons: legal name, plus trading name(s) and acronym(s).

5.18.   For natural persons: family name(s), names normally used and possible pseudonyms, forenames, personal number (if authorized by national law), gender.

5.19.   For statistical units: trading name(s) and acronym(s).

5.20.   For administrative units: names as registered.

Address

5.21.   Physical address, including postal code and official geographical code at the most detailed level.

Communication characteristics[35]

> ‣   Telephone and fax number(s)
> ‣   Electronic mail address(es)
> ‣   Information on sources of quick updates about the unit characteristics (such as websites)
> ‣   Mailing address (if different from the legal one)

5.22.   *Purpose:* These characteristics provide contact information for the unit. Telephone number and email address are very important. In the event that units from different sources do not share a common identification system, their names and addresses can be used for matching units across sources.

5.3   Demographic characteristics[36]

Date of incorporation of unit (for units that are legal persons)

Date of official recognition as an economic operator (for units that are natural persons)

5.23.   *Purpose:* One or other of these characteristics should be recorded at the time of inclusion of a new unit, whether this is the result of real birth or of another demographic event that results in a new unit, for example the incorporation of an unincorporated enterprise.

5.24.   *Definition:* The date of official recognition may be the date on which an identification number is assigned, or the date in which the legal existence is accepted, be it through a trade register number, the assignment of a VAT number, or other form of administrative registration.

Date of commencement of economic activity

5.25.   *Purpose:* The characteristic is used in deciding if a new unit is economically active.

5.26.   *Definition:* The date of commencement of economic activity should be the date on which the unit actually starts its economic activity.

5.27.   As soon as a unit invests it is deemed to have economic activity even though production may not have started. If this were not the case, then investments, in particular gross fixed capital formation, could be heavily underestimated.

5.28.   The actual date of commencement of economic activity is often difficult to know. A proxy for this date can be the date declared in advance by the unit at the time it is officially registered and an identification number is assigned. Another proxy is the date when its legal existence is accepted.

Date of final cessation of economic activity

Date on which a legal unit ceased to exist

5.29.   *Purpose:* These characteristics are needed to record the permanent inactivity or death of a unit. Death is a difficult status to establish with precision as liquidation can last a long time after economic activities have been stopped, in particular when lawsuits are in process.

[35] The communication characteristics of the unit may be different for different surveys. They may include the names and addresses of the corresponding contact persons. In addition, as previously noted, the contact persons might be reporting units such as accounting or tax consultants hired by the unit.

[36] *Eurostat-OECD Manual on Business Demography Statistics* (Paris, 2008, and the ECE *Guidelines on the Use of Statistical Business Registers for Business Demography and Entrepreneurship Statistics.* Geneva, 2018

Date of merger, takeover, split or break

Identifier of unit (if any) that is continuing the economic activity after the concerned unit has ceased to exist

5.30.   *Purpose:* These demographic characteristics (further discussed in sections 7.3 and 9.3) are needed to record the restructuring of units and to relate the resulting live units to the previous ones that have been ceased and might be considered dead.

5.4   Economic stratification characteristics

Legal form

5.31.   *Purpose:* The legal form (also known as legal status) is a key stratification characteristic. It is important to avoid ambiguity and double counting in identification searches. The legal form of a legal unit is useful for the stratification for surveys and the classification in the institutional sector.

5.32.   *Definition:* The possible types of legal form differ across countries according to the particular legislation in each country. For unincorporated enterprises there are different types of legal forms, including sole proprietorship and partnership. A typical set of legal forms for European Union countries is listed in section 4.7.1.

5.33.   While the range of legal forms varies from country to country, depending upon national legislation, there are some legal forms that can be found in most countries.[37]

5.34.   The tax regime applicable to a legal unit depends on its legal form, which means that taxation data supplied to an SBR depend on legal form. Some legal forms may be exempted from certain regulations (VAT, for example), in which case the corresponding legal units are not registered by the corresponding administrative sources, adversely affecting the coverage provided by these sources.

Institutional sector and subsector

5.35.   *Purpose:* The institutional sector classification has at least five important functions:

▸   In national accounts, for compiling data for individual sectors.

▸   In business statistics and business demography, for separating market and non-market activities.

▸   In separating data for public enterprises from data for private enterprises.

▸   In finding the share of employment, or value added, in a country's economy of which the country itself is the ultimate controller.

▸   In distinguishing foreign controlled enterprises.

5.36.   The set of institutional unit types is described in chapter 3, and the institutional sector classification is presented in annex B2.

Principal economic activity

5.37.   *Purpose:* The principal economic activity code of a unit is a key stratification characteristic.

5.38.   The activity code is based on a classification of economic activities. The recommended international classification for economic activities is ISIC Rev. 4, presented in annex B1.

[37] As indicated in the glossary of the *Eurostat-OECD Manual on Business Demography Statistics,* 2008..

Secondary economic activity (activities)

5.39.   Practical guidance in assigning and maintaining principal and secondary economic activity codes is provided in chapter 7.

Ancillary unit

5.40.   *Purpose:* Some units perform activities exclusively for other units without autonomy. As a consequence, they have no economic significance per se. They should be seen as ancillary activities of the parent legal unit that they serve, to which they belong and to which they must be attached to form a unit significant for economic analysis.

Market orientation (market, non-market)

5.41.   *Purpose:* Both the market- and non-market characteristic of a unit and its legal form are needed to classify a statistical unit to the appropriate institutional sector.

Turnover

5.42.   The meaning of the terminology used in business accounting may vary greatly from one country to another and the term "turnover" may have different meanings. Here the terms "turnover", "sales" and "revenues" are used interchangeably.[38]

Flag indicating consolidated turnover

5.43.   *Purpose:* The SBR should record an estimate of economic size for every unit. For production statistics, value added is considered the most appropriate characteristic, but it is difficult to measure and is often replaced by proxy measures. Because it is relatively easy to collect, the most commonly used monetary proxy is turnover (revenue or sales). A commonly used physical proxy is number of persons employed.

5.44.   There are a number of drawbacks to using turnover, due to the need for consolidation of internal flows between the statistical or reporting units:

> ‣   Turnover is not additive when changing the level of observation units, for example from establishment to enterprise, or from enterprise to enterprise group, or from the usual observation unit (which is typically a legal unit) to enterprise or enterprise group.

> ‣   Turnover is not additive when changing the geographical level; it is different at global enterprise group level than at truncated national enterprise group level.

5.45.   There are different definitions of turnover according to the economic activity of the unit being classified. For example, there are differing definitions for agriculture, forestry and fishing, manufacturing, finance and insurance, and economic activities of households as employers.

5.46.   If the consolidated turnover of a statistical unit is not available it is better to use as proxy the sum of the turnovers of its constituent units than to have no measurement at all. As turnover of legal units is used in most administrative sources, it should be recorded for consistency purposes.

[38] For a more detailed description of "turnover" reference is made to para 4.144 and table IV.1 of the International Recommendations for Industrial Statistics 2008 (United Nations, 2009).

Number of persons employed

Number of employees

5.47.   For all units the SBR should record the actual number of persons employed, and number of employees, both as head counts and, in the latter case, also in full-time equivalents (FTEs). FTE employment is defined as total hours actually worked by all employed persons divided by the average number of hours actually worked[39] in full-time jobs (2008 SNA, para. 19.43).

5.48.   *Purpose:* The main uses of these characteristics are in stratification for sampling, analysis and dissemination purposes.

> ‣   Number of persons employed is preferable for stratifying survey samples in the case of very small units.

> ‣   Depending on the source and quality of the information, employment data directly from the SBR may be used in compiling employment statistics. They are especially useful for small area statistics, where the SBR may be the only comprehensive source.

5.49.   "Number of persons employed" is defined as the total number of persons who work in the unit, including wage earners and self-employed persons (namely, working proprietors, partners working regularly in the unit and unpaid family workers), and also persons who work outside the unit but who belong to it and are paid by it (such as sales representatives, delivery personnel, repair and maintenance teams). It excludes labour supplied to the unit by other enterprises, persons carrying out repair and maintenance work in the unit on behalf of other enterprises, and those on compulsory military service.

5.50.   "Number of employees" is defined as those persons who work for an employer and who have a contract of employment and receive compensation in the form of wages, salaries, fees, gratuities, piecework pay or remuneration in kind. A worker from an employment agency is considered to be an employee of that employment agency and not of the units in which he or she works, including temporarily.

5.51.   For stratification purposes the number of employees at year end or the annual average number of employees may be used. An annual average can also be calculated over a specified period within a year for seasonally active units.

5.52.   The number of employees in FTEs may be calculated for the period during which the unit is active or for a full year. For stratification purposes the active period should be used while for annual economic statistics the whole year is more appropriate.

5.53.   *Proxies*: In some countries all these data can be obtained directly from administrative sources while in other countries administrative sources may provide only the number of paid employees. In the latter case, the number of persons employed can be estimated according to legal form and activity. For example:

> ‣   For sole proprietors, number of persons employed = number of employees+1.

> ‣   For two-person partnerships, number of persons employed = number of employees+2.

5.54.   *Sources:* Administrative sources, surveys and SBR calculations.

5.55.   *Comments:* The reference period used for the measurement of employment in business demography is a calendar year. Thus, in principle, a labour force estimate should be an annual average. If a unit operates only during a part of the year (as in the case of a seasonal enterprise or a new enterprise), the average for size measurement

[39]   For a definition of "hours actually worked", reference is made to the 2008 SNA, paragraphs 19.51–19.52 and, for a detailed discussion of the criteria for inclusion and exclusion in "hours actually worked", to paragraph 11 of resolution 1 of the Eighteenth International Conference of Labour Statisticians, in 2008.

should be based on that period. For annual employment data, however, the average should be calculated for the whole calendar year. How the annual average is calculated in practice depends on the updating frequency of the SBR and on the sources used.[40] For example, the value may be approximated by using the number of persons employed at a specified moment during the year if this is all that is available.

5.56.   Head counts and FTEs each have certain advantages, so both should be recorded if possible. Head count is the number of physical persons employed by a unit, full-time and part-time. FTE is a more accurate measure of labour input but, depending upon the origin of the data, it may not be possible to calculate. Furthermore, as the concept of "full-time" may vary from one country to another, the FTE data may not be fully comparable across countries. Another possibility is to use "hours actually worked" directly as a measure of work input. This measurement is gaining favour in employment statistics, but data are not often readily available.

5.5   Relationships and links between units in the SBR

5.57.   There are two main reasons for registering relationships between the various types of units in the SBR:

> ‣   For management and the direct use in the SBR itself and by surveys using frames derived from the SBR, relationships are required to link together units that belong to the same enterprise or to the same enterprise group, and to ensure correct transformation of data collected from administrative units, or about observation units, to data for the corresponding statistical units.

> ‣   For potential uses of the SBR, relationships assist in statistical and economic studies that depend upon bringing data from different sources together.

5.58.   The most evident need concerns two types of links: those between statistical units and administrative units required for creation and management of the SBR, and those between statistical units and observation units. Depending on the sources used in building the SBR and on its main uses, there may also be a need to register links between administrative units and legal units, and between observation units and reporting units.

5.59.   Each link should include the date when the relationship started and, where applicable, the date when it stopped.

5.60.   In its simplest form a relationship is a link including only two identification numbers (one for each unit linked) and the date. In some situations more complex relationships are recorded. (For example, a single unit of one type of unit may be linked to *n* units of another type, and the relationship may indicate the percentages to be applied in splitting data for the one unit across the n units.) This depends on how the SBR is organized. In general, it is best to keep the relationships as simple as possible.

5.61.   It is important to recognize the difference between the treatment of administrative registers and the treatment of the SBR. In an administrative register, units and their values may be changed without notice to the national statistical office and without continuity from a statistical perspective. It is thus essential to keep copies of the administrative registers used as sources by the SBR and, where possible and relevant, to record the relationships between the units in these registers and legal units and statistical units.

[40] Fiscal sources are related to the country's fiscal year which can be different from the calendar year. Their average employment then refers to a different period.

5.62.   Links should also be kept between the observation units and the corresponding reporting units if this is practical for survey purposes. For example, in some cases a single lawyer or accountant may report on behalf of several enterprises. These links can be maintained in the SBR or in survey specific data-collection systems.

5.6   Relationships and links with other registers

5.63.   Whether the SBR contains data for all the different unit types in a single database, or in separate satellite registers, such links are very important, given the emerging role of the SBR in connecting various data sources and thus integrating statistics.

5.64.   Links are expressed in terms of identification numbers for the external sources and the identification numbers of the corresponding units in the SBR.

5.65.   Ideally, the links should be organized in such a way that all administrative units are linked to the legal unit to which they belong. For example, a legal unit may have several administrative units for pay-as-you-earn purposes, and several more for social security contributions, in addition to having one or more VAT accounts and an account for payment of income tax.

5.66.   Depending on national legislation, it is conceivable that a single administrative unit is linked to more than one legal unit.

5.67.   The sorts of relationships that exist and links that are possible depend upon country legislation and regulations. Examples are:

>   Links from the legal unit to VAT
>   Links from the legal unit to the trade register
>   Links from the legal unit to social security
>   Links to balance sheet data (for units required to publish accounts)
>   Links to the balance of payments register or FDI register
>   Links to the farm register
>   Links to other satellite registers, for example, shop registers, tourist establishment registers, transport registers, educational services registers, and health services registers

5.7   Characteristics by unit type

5.68.   Figure 5.1 summarizes how the characteristics described in general terms in the previous sections are applied to each of the types of statistical units defined in chapter 4. In the table a tick mark (✓) indicates which characteristics should be recorded for the different types of units listed in the columns. A blank cell indicates that the characteristic is not applicable for logical or conceptual reasons. For instance, the characteristic "type of enterprise group" is only applicable for the statistical unit enterprise group, but not for the other units.

5.69.   Annex A provides comprehensive details about the characteristics of the units by type of unit.

Figure 5.1
Summary of characteristics by unit type

Characteristics of the unit	Enterprise group and truncated enterprise group	Enterprise	Establishment	Local unit (of enterprise)	Legal unit	Other observation unit (VAT, social security units, etc.)
Identification and contact						
Type of unit	✓	✓	✓	✓	✓	✓
Type of enterprise group	✓					
ID number	✓	✓	✓	✓	✓	✓
Name, address	✓	✓	✓	✓	✓	
Communication characteristics, contact details	✓	✓	✓	✓	✓	✓
Administrative identification numbers (tax number…)					✓	
Geographical code (postal code…)	✓	✓	✓	✓	✓	✓
Country code	✓	✓	✓	✓	✓	✓
Demographic						
Birth date	✓	✓	✓	✓	✓	
Status of unit (active/inactive….)	✓	✓	✓	✓	✓	
Restructuring code (merger, split…)	✓	✓			✓	
Restructuring date	✓	✓			✓	
Death date	✓	✓	✓	✓	✓	
Economic stratification						
Legal form					✓	
Institutional sector		✓			✓	
Principal activity	✓	✓	✓	✓	✓	
Secondary activity (if any)	✓	✓	✓	✓	✓	
Market oriented		✓	✓	✓	✓	✓
Ancillary unit		✓	✓	✓	✓	
Employment	✓	✓	✓	✓	✓	
Consolidated turnover	✓	✓	✓			
Links and external references						
Link between unit and enterprise group		✓			✓	✓
Percentage control					✓	
Percentage share (ownership)					✓	
Type of management/control unit (within group)					✓	✓
Link between unit and enterprise			✓	✓	✓	✓
Link between unit and establishment		✓		✓	✓	✓
Link between unit and local unit (of enterprise)		✓	✓		✓	
Link between unit and legal unit	✓	✓				✓
Links between units involved in mergers, splits, etc.	✓	✓	✓	✓	✓	✓
External register - name of register, identification number(s)			✓		✓	✓

Chapter 6
Data sources for the statistical business register

6.1 Introduction

Context

6.1. This chapter provides guidance on the sources of data that can be used to construct an SBR and how they are used in combination. The main focus is on administrative data, but statistical sources and new potential sources are also discussed. The chapter defines what is meant by "administrative data" and describes main administrative sources that are used in many countries. The advantages and disadvantages of these sources are considered, and also the legal aspects pertinent to their use.

6.2. Different data sources may be used in combination to construct and maintain an SBR with good coverage and content. The data sources must also provide information about changes in units to keep coverage and content up to date, as further discussed in chapter 7. Unless the SBR is maintained on a regular basis, it quickly loses its value by becoming outdated and ceasing to adequately reflect economic activity in the real world.

6.3. Data sources may be grouped into three categories:

› Administrative sources, such as business registration and licence register, tax register, company and trade associations and chambers of commerce registers, social security registers and others.

› Statistical sources: economic census and agricultural census, feedback from economic surveys, profiling and SBR improvement or maintenance surveys,

› Other sources, for example data from private data suppliers, telephone directories and the Internet.

6.4. The Guidelines recommend that SBRs be created and maintained primarily using administrative sources. Typically, they enable good coverage and stability. This approach is in line with Principle 5 of the United Nations Fundamental Principles of Official Statistics, which states:

> 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.

6.5. While in many countries administrative data are the main sources for the updating of their SBRs, there are also countries where the national statistical office does not have access to administrative data, comprehensive administrative data do not exist, or the administrative data are of insufficient quality for the updating of the

SBRs. In these countries the traditional economic censuses or other kinds of business surveys, together with household surveys, are therefore an indispensable source for the SBRs. Economic censuses, mostly based on the establishment unit, deliver relevant and core information for the SBR on each single establishment, such as address, economic activity, legal unit, number of employees and others. Even in countries that are able to use administrative data for their SBRs, data from censuses may be used as a complementary source for checking or updating the SBR.

6.6.   In general, the best choice of data sources depends on the specific situation in any given country, including the availability of administrative data and the scope and complexity of the national statistical system itself.

Administrative sources

6.7.   The term "administrative sources" is commonly understood to mean sources containing information that is not primarily collected for statistical purposes. Another European Union source[41] defines administrative data as: "data derived from an administrative source, before any processing or validation by the national statistical office". These definitions allow for the wide interpretation of the data that might be included under the heading "administrative data" in these Guidelines. A more traditional and narrower definition requires administrative data to be collected by government bodies for the purpose of administering taxes, public pension funds and other regulations.

6.8.   Eurostat collects information on administrative and other data sources used for SBRs in its annual business registers questionnaire. The same questionnaire is also used by the ECE and OECD and thus covers a wide range of countries. The most commonly used sources relate to taxation systems, such as VAT and personal income tax, and to compulsory business registration systems, often administered by chambers of commerce or government bodies and departments.

Statistical sources and survey feedback

6.9.   Statistical sources refer to data collected by statistical processes carried out by the national statistical office and, in principle, by other producers of statistics within the national statistical system. The three survey groups of sources are, first, the economic surveys that draw their frames from the SBR and that can provide updates; second, improvement surveys conducted by the SBR itself, and, third, economic censuses which provide comprehensive data that can be used to develop and update the SBR. Another very important source is profiling, a method of analysing the legal, operational and accounting structure of large and complex enterprise groups, in order to establish the statistical units, their links and the most efficient structures for the collection of statistical data. Profiling may be carried out at a purely national level or approached from a more global and collaborative perspective, as in the methodology of European profiling.

6.10.   Economic censuses usually exclude the agricultural sector. Business units active in the agricultural sector (crop and animal production, hunting, forestry, and fishing) still need to be covered in the SBR, however. Accordingly, for many countries the data from the agricultural censuses are equally important for the maintenance of the SBR as the economic censuses. In the box below some examples are given.

[41] From ESSnet on Administrative data, a European Union-funded project.

Box 6.1
Examples of uses of economic censuses in the SBR: China, Egypt, Indonesia and the State of Palestine

China

China conducts an economic census every five years. The economic census is very important for understanding the structure and development trend of economic entities. It also offers an effective and comprehensive means for developing countries to observe economic growth. Between the economic census years, the maintenance of the basic unit registry is usually divided into two parts for easier management: the units above a designated size will be reviewed and adjusted on a monthly basis; for other smaller units, the changes and updates will be submitted and reviewed on a quarterly basis.

Egypt

In Egypt, the economic census provides important updates to the SBR, such as contact details, main and secondary economic activities, number of employees, start-up date of the activity and death of the unit, and the business profile, which represents the structure of the enterprise groups.

Indonesia

Based on the results of a study on administrative data in Indonesia, tax data were considered the best administrative source for updating the SBR. Because the confidentiality of taxpayer data was protected by law, however, it had not been possible for BPS to gain direct access to taxpayer data until 2018. There were other administrative sources with good coverage, such as data from a business licensing agency, but they were not as complete as tax data and usually were not updated frequently. As a result, there were no reliable administrative data for updating the SBR, and BPS used the economic census results for that purpose, in addition to profiling results and feedback from surveys that used the SBR as the sampling frame. After a successful negotiation with the Tax Office, however, BPS will be able to use tax data as a source of SBR in the future.

State of Palestine

As of 2017, establishment censuses are still the main source of data and are considered a framework for the implementation of specialized surveys in which the establishment is the sampling unit.

In the State of Palestine, a census of the establishments is carried out every five years. The establishments are the main sampling unit and a limited update for the establishment framework is carried out between each census, owing to the lack of comprehensive periodic updating sources. The lack of comprehensive updating of the sampling frame weakens the content of the frame as a result of the rapid changes occurring in the actual economic situation, which are not addressed by the framework in a regular and comprehensive manner. The following are examples of such changes:

1. New establishments entering the market are not included in the sampling frame.
2. Establishments cease their economic activity while their entries are not removed from the statistical framework.
3. Structural changes are made in some establishments, such as changes in their economic activities or changes in their economic size.
4. Registered establishments without a fixed place to carry out their activities are not covered in the establishment census.

These problems are evident when the results of a new establishment census are compared with the results of the previous one. Newly created establishments are not covered by the SBR and are thus not included in the survey frames, whereas ceased establishments are still covered in the frames. PCBS overcomes this problem by using data from the

Box 6.1 (*continued*)

Quarterly Labour Force Survey to update the establishments in the labour market between the censuses, while there is still a lack of coverage of the establishments that are listed in the administrative registers, but do not have a fixed location from which to exercise their activities.

PCBS has started to focus on administrative data sources to construct and update the SBR, and the results of the 2017 population, housing and establishment census were taken into account as the backbone. The procedures that are being worked out in this field are summarized in the collection and diagnosis of administrative data from several sources related to the system of registering establishments, dealing with legal aspects of the use of administrative data in the structure of the registry, linking, comparing and processing administrative data with census data, studying the mechanisms and methods of linking administrative data with census data to identify the same establishment from both sources by comparing the identification characteristics of the establishment, arising from the data of administrative records.

The experience of the State of Palestine in this situation is still under development, owing to the incompleteness of the administrative records. PCBS is endeavouring to follow a number of different methods and procedures aimed at integrating the sources of census data and administrative records to be used to build and update the SBR, including cooperation with partners to ensure that data from different sources are updated and linked to the SBR updates. PCBS also establishes a common identification system to facilitate the work of combining data from various administrative sources, and to identify and use the most useful administrative data sources in terms of quality and coverage to construct and develop the SBR.

Box 6.2
Examples of uses of agricultural censuses in statistical business registers: China, Colombia, Indonesia and the State of Palestine

China

In addition to conducting economic censuses, the National Bureau of Statistics also carries out an agricultural census every 10 years. The agricultural census is a comprehensive survey of farmers, agricultural production, and agricultural units. Results from the agricultural census throw light on developments and changes in agriculture, rural areas and farmers in a comprehensive way, in particular at a time when the continuing agricultural reform and development in China has entered a new era. The Bureau is also facing new challenges in conducting agricultural censuses, such as restrictions on the organization of face-to-face interviews and difficulties in capturing technological changes in the shift from traditional to modern agriculture. The Bureau is promoting the use of remote sensing measurement, hand-held intelligent data-acquisition terminals, rural census data networking for direct reporting, and other modern information technology that could facilitate data collection during the agricultural survey.

Colombia

Agricultural production units in Colombia are determined by satellite cartography and confronted with the information from the country's cadastral authority – the Agustín Codazzi Geographical Institute. The updating of these units is accomplished through the agricultural census, national agricultural survey, and information provided by guilds and large producers.

In different sectors, different data sources are used, such as production guilds. The commercial data available from private companies that manage public services, such as telephone companies and water and energy utilities, are used in national accounts to complement or validate the primary data obtained through surveys.

Indonesia

Indonesia holds an agricultural census every 10 years. The types of units enumerated in the census are:

- Agricultural companies (corporations)
- Households
- Others (not corporations or households), such as Islamic boarding schools or orphanages that also carry out agricultural activities

In the period between censuses, BPS updates the agricultural company data every year by conducting surveys arranged by their respective subject matter. Meanwhile, BPS updates agricultural household data with surveys every three to five years.

Besides collecting data in the field, BPS also matches the agricultural directory data with administrative data from the Ministry of Agriculture and agricultural associations. The agricultural data from the Ministry of Agriculture include lists of agricultural businesses that have permits, but these are not updated if the business has closed or moved to a new location. Meanwhile, farm household businesses usually do not have permits, but most of them are members of farmer groups registered in the agricultural department in the region. In the agricultural household survey, there is a question about farmers group which the farmers have joined.

State of Palestine

Before an agricultural census is carried out, the available data on the agricultural situation in the State of Palestine are estimated by the Ministry of Agriculture, based on their locality and not on the agricultural holdings according to FAO recommendations. The First agricultural census was conducted in 2010 by PCBS and the Ministry of Agriculture. Results from that 2010 census indicated that most of the agricultural holdings were held by individuals and households (99.9 per cent of agricultural holdings), and 75.5 per cent of agricultural holdings were classified as small (less than 1 hectare).

In general, PCBS is trying to use administrative data as much as possible for the production of annual statistics from line ministries, including the Ministry of Agriculture. PCBS requested the Ministry of Agriculture to update annually the agricultural census frame, using complete enumeration. This process proved unsuccessful, owing to the lack of human and financial resources to complete the task and the low turnout by farmers to come to the locality offices to update their data.

Currently, the Ministry of Agriculture is planning to set up a registry system for agricultural holdings and agricultural facilities in the State of Palestine that will include all basic information about the holdings, based on data from the agricultural census, together with other necessary information about the status of the agricultural sector. A web-based application is also being designed for data acquisition to facilitate data collection and verification. In addition, PCBS and the Ministry of Agriculture are planning to conduct the second agricultural census in 2020.

Commercial and other data sources

6.11.   The Guidelines consider that commercial data available from private companies that manage utilities, for example, telephone companies, water or power supply, may be considered in the same way as administrative data. Use of such data can be a cost-effective way of improving the SBR, since these databases are extensively researched and managed for commercial benefit by their owners.

6.12.   Other sources include data collected by companies for their own commercial purposes, for example data owned by market information companies and companies that collect data and maintain directories and databases that they provide to their customers for commercial purposes.

6.13.   New data sources are emerging and current discussion of big data may well highlight additional sources. Such sources may be used in SBRs, for example in helping to create enterprise groups.

Content of chapter

6.14.   In the following sections each of the various types of sources is discussed in more detail and the final section describes the record linkage procedures that are required to combine data from them, in case common unique identifiers are not given. Influential reference documents include the Eurostat *Business Registers Recommendations Manual* and the African Development Bank's *Guidelines for Building Statistical Business Registers in Africa*.

6.2   General methods, procedures and issues

6.2.1   Cooperation with data providers

6.15.   Establishing and maintaining good relations with the owners of data sources, in particular administrative sources, is vital, and can be complex and challenging. From an SBR perspective the purposes of cooperation are to understand the concepts, to ensure continuity of supply, and to ensure easy linkage of data from the various sources to the SBR, preferably through a common identifier. In other words, the aim is to promote the interests of the SBR and, more generally, of the economic statistics programme. The purposes are the same, whether the SBR is dealing with an administrative, statistical or commercial data source.

6.16.   Although cooperation can take on many forms, it is recommended that it include regular face-to-face contact and mutual visits. A strong commitment from the senior management of the national statistical office and leaders of the administrative data providers is required, and meetings at the managerial level are crucial in order to actualize the desired cooperation. Contact should involve both management level staff and operational staff who are working with the data on a daily basis. Establishing a mechanism for regular consultation would also be important to monitor progress in the development of an SBR.

6.17.   The SBR unit can also play a central role within the national statistical office in the general coordination with the administrative data providers. Administrative data required for the SBR are often also important data for survey and other statistical purposes (such as employment data and turnover data, for example). It could be a task of the SBR unit to coordinate the various administrative data requirements within the national statistical office and communicate it with the respective administrative data providers. The SBR unit, if adequately staffed, can also take care of the data delivery,

storage and quality control of the administrative data received. In some cases, it might even be necessary to establish a separate team for administrative data acquisition.

6.18.   Through effective cooperation it is possible to be proactive regarding changes in data sources that might be detrimental to the SBR, and to limit undesirable changes, or at least to have early warning of upcoming changes that could disrupt, jeopardize or otherwise challenge statistical production or accuracy. Cooperation should also be used to promote the benefits of using statistical concepts in administrative sources, for example, standard economic activity codes and standard definitions of business units. These considerations should ideally be made when new administrative data sets are being established.

6.19.   To provide clear responsibility and working arrangements it is good practice to set up a memorandum of understanding or service-level agreement with an administrative source. While not often legally binding, a memorandum of understanding or a service-level agreement provides an agreed framework and gives reassurance about the services to be received. Typically, a service-level agreement sets out the agreed data coverage and content to be supplied, delivery timetables, data security provisions, checking mechanisms and quality provisions, and response times within which to answer validation questions by the SBR.

6.2.2   Common identification systems

6.20.   A common identification system eases the work in combining data from administrative sources. If there is no common administrative identification system, the national statistical office should use any opportunity to promote one and to draw attention to the advantages of linking data in terms of more accurate and less costly statistics.

6.21.   Linking different data sources is not only a means of improving the quality of statistics but also of reducing administration costs and burden. If a common administrative identification system does not exist, the national statistical office should construct an internal linkage table (containing links between units in the various sources) in order to manage the SBR and to reduce duplication or omission problems.

Box 6.3

The legal entity identifier and the Global Legal Entity Identifier Foundation

The legal entity identifier (LEI) provides clear and unique identification for each legal entity participating in financial transactions across the globe. LEI codes are centrally managed by the Global Legal Entity Identifier Foundation (GLEIF), a not-for-profit organization headquartered in Basel, Switzerland, and issued by local operating units that are accredited by GLEIF. Each LEI uniquely identifies a given legal entity that is legally or financially responsible for the performance of financial transactions, such as a corporation, trust, partnership, and so on. LEI remains the global industry standard best suited to providing open and reliable data for unique legal entity identification management.

In 2011, the Group of Twenty called on the Financial Stability Board to provide recommendations for a global LEI and a supporting governance structure. In response, the Board created GLEIF in June 2014 and entrusted it with supporting the implementation and use of the LEI. GLEIF is backed and overseen by the LEI Regulatory Oversight Committee, which represents public authorities from around the globe that have come together to jointly drive forward transparency within the global financial markets. One of the services that GLEIF provides is the Global LEI Index, which contains

> **Box 6.3** (*continued*)
>
> historical and current LEI records including related reference data in a single authoritative, central repository. It is the only global online source that provides open, standardized and high-quality legal entity reference data. Any interested party can easily access and search the complete LEI data pool free of charge on the GLEIF website using the web-based LEI search tool developed by GLEIF.
>
> Public authorities are increasingly relying on LEI for their regulatory purposes. For example, a new European Union rule (MiFID II/MIFIR) that took effect on 3 January 2018 requires investment firms to obtain LEIs from their clients before providing services which would trigger related reporting obligations.
>
> More information on LEI data and GLEIF services may be found on the GLEIF website (www.gleif.org/).

6.2.3   Identifying and using external data sources

6.22.   The first step in using administrative data is to identify the most useful data source or sources. Potential sources have to be evaluated in terms of their coverage, content and the costs to the SBR of acquiring the data. For evaluation purposes it is important to obtain all available information about each source, for example, the definitions that are used, the coverage, the updating methods, the frequency of updates, the time lag, and how frequently it is possible to get information from the source.

6.23.   Data obtained from external sources should be stored in the SBR without change of content. Retaining the original data ensures that there is a record of what was received, regardless of changes made during subsequent administrative source processing or by SBR updating processes. This is important for audit purposes and for communication with the administrative sources should there be any issues or concerns about changes that have been made. It also enables the SBR to be rebuilt if any system or processing issues occur during production and updating processes. Administrative data are often extracted from live databases and cannot be replicated for the same reference period at a later date, so keeping the administrative data in its original state is good practice.

6.24.   Storing metadata about changes in data values is essential. For example, if a unit changes address in an administrative register, the old address should be retained in the SBR but with a termination date indicating that it is no longer current and the new address should be stored with the commencement date and a blank termination date.

6.25.   Following receipt and storage of administrative data, statistical units in the SBR are created or updated through a set of conversion rules and procedures, as further described below.

6.2.4   Combined administrative and statistical register

6.26.   Some countries have developed a single administrative business register with multiple functions, including that of an SBR. This has significant advantages in that statistical data requirements are within the core of the administrative systems. Country examples from Canada, Costa Rica, Denmark, Georgia, Indonesia and Malaysia on the integration of administrative registers and SBRs are presented in annex C.

6.2.5   Linking and matching

6.27.   The SBR is more resilient with lower risk of error, if multiple reliable data sources are used. Some sources, such as VAT, provide extensive (but never complete) coverage of the economy, whereas others may only cover one sector, such as the financial sector, but more completely. Thus, there are significant advantages to using data from several sources. The use of two or more data sources, however, poses two types of linking and matching challenges:

> ‣   First, where the sources have a similar (and thus duplicated) coverage of the economy, the challenge is avoiding duplication without incurring omission.

> ‣   Second, where each source covers only a limited sector of the economy, such as finance and public administration, the challenge is dealing with the sector boundaries associated with each source.

6.28.   If there is no common identification code, a probabilistic approach to linking can be adopted, based on similarities in the name, address, telephone and fax number, or other characteristics, such as legal form and economic activity code, of the units in the different sources. This process is generally referred to as "matching". It can lead to units being linked in error, however, resulting in "mismatches" (also called "false" or "incorrect matches"), and in units not being linked, or "missed matches", which result in duplication. Experience over the years in several national statistical offices suggests that linking large data sets in the absence of a common identification code is difficult and requires substantial investment in software and systems. Matching is further discussed in section 6.8.

6.2.6   Integration of the administrative units into the SBR

6.29.   After administrative data have been acquired, the next step is to match them with the administrative units that are already covered in the SBR. This is relatively easy if a common identification number is available, and there are no errors in this number in either the administrative sources or the SBR. There are occasions, however, where an administrative unit is different over the various administrative sources.

6.30.   The matching results should be quality checked before entering the database. An identical unit from the SBR and the administrative source may not match as they may not have the same exact values in both databases (different updating procedures, errors in the data). Checking may involve comparison of the values of characteristics such as name, address, economic activity code, size or legal form across the linked units. If these values are consistent, it is more likely that the link is correct. If not, further clerical checks may be needed, particularly where larger units are concerned. If additional data, such as turnover and import and export data, are available from other administrative sources, these can also be used in the checking procedure. Checking might also involve contacting the unit.

6.31.   It is also useful to periodically check administrative units that have not been matched and attempt to establish further links or to determine why they do not match. If the non-matched units do represent active economic units, the failure to match may be due to timing or scope differences between the administrative source and the SBR.

6.2.7   Transforming administrative units into statistical units

6.32.   The final step is to build the statistical units, mainly the enterprise, on the basis of the administrative and legal units. In the normal case an enterprise coincides with a legal unit. This is the case when the legal unit is not controlled by another legal unit and thus has autonomy. Most enterprises belong to this category.

6.33.   In cases, however, where a legal unit is controlled by another legal unit, this unit may be viewed as an enterprise only if it has autonomy. Thus, legal units that belong to an enterprise group may not necessarily be considered as an enterprise and may need to be combined with another legal unit of that group to form an enterprise. Such cases are normally quite complex to handle and need usually some investigation, either using published information of that group or direct contact with the group head or decision centre in the country. The whole task of delineating of the statistical units based on legal units is called "profiling" and is further described in section 6.6.4 (for the definition of the statistical units see chap. 4).

6.3   Administrative data sources

6.3.1   Typical administrative sources

6.34.   Some of the commonly used data sources are listed below. The list is not complete; other data sources may be available.

> ▸   *Business registration and licence register*: In some countries it is compulsory to register or license a business before trading. If available, this register can provide basic information on identification number, name, address and other contact information. This sort of register may be run by the tax authority, Chamber of Commerce, licensing office, or another public authority.

> ▸   *Tax registers*: Typically relating to VAT or employee income tax, tax registers may be a source of data on economic activity, turnover, and activity status.

> ▸   *Company and trade associations and chambers of commerce registers*: These can provide information on economic activity, legal form, and births and deaths.

> ▸   *Social security registers*: For businesses employing paid staff and making social contributions for employees, a social security register can provide identifying characteristics and stratification characteristics, such as legal form and number of employees.

> ▸   *Labour and employment registers*: Such registers can provide additional economic and social information about employees.

> ▸   *Government units registers*: Such registers are maintained by government finance departments for financial management of the public sector.

> ▸   *Non-profit unit registers*: Such registers are maintained by regulators, for example of charities, and typically contain name, address, other contact information, legal status and possibly economic activity code.

> ▸   *Industry association registers*: Such registers may contain name, address, other contact information and economic activity code. They are likely to be up-to-date, but only contain members of the association, so completeness may be an issue.

> ▸ *Agricultural administrative registers*: Such registers may cover agricultural holdings as distinct from businesses. They typically contain name, address, other contact information and indicators of economic activity and possibly size.

> ▸ *Water supply and electric association registers*: Such registers are maintained by public or private utility bodies. They typically contain name, address, other contact information and indicators of economic activity and possibly size.

> ▸ *Sector-specific source*: Sources include lists of schools from the education ministry, lists of hospitals from the health ministry and lists of charities from regulators. While coverage is limited to a specific sector, within that sector it may be very comprehensive.

> ▸ *Central banks*: Central banks often have information for the financial sector and on units engaged in FDIs, from supervisory authorities.

> ▸ *Published business account*: Data from financial reports are particularly valuable as they contain information on shareholders and subsidiaries that is essential in delineating enterprise groups. National statistical offices are encouraged to explore the possibilities of automatically extracting data from internal financial or management accounting systems of businesses, for example, using XBRL.

6.35.   Besides providing names and contact information, these sources may also contain data that indicate whether or not a unit is active, its principal economic activity, its size and some other variables relevant for the SBR, such as employment and turnover.

Box 6.4
Examples of the use of tax data in the SBR: Colombia, Mauritius, Portugal, South Africa and Tunisia

Colombia

The Directorate of Taxes and Customs is a relevant source of data for the SBR, considering the degree of coverage and reliability of the information. The regulations that restrict its disclosure and use mean, however, that access to these records is limited.

It is possible to access specific attributes through the definition of specific agreements for the exchange of information, supported by technical committees and with the corresponding legal support. Thus, the SBR has included this data source by considerably expanding the stored records and allowing to carry out processes to improve the quality and consistency of the data. Currently, the National Administrative Department of Statistics seeks to include the Tax Office as a permanent source of information, with defined exchange protocols and not just as an occasional source.

Mauritius

Consultation between Statistics Mauritius and the Mauritius Revenue Authority (MRA) started in 2013. Since then several working sessions, meetings, presentations, including meetings at directorate level (Statistics Mauritius and MRA) and the Ministry of Finance (parent Ministry of Statistics Mauritius and MRA) have been conducted. At the present time, data for corporations have been provided to Statistics Mauritius by MRA, and discussions are still ongoing to acquire data for small enterprises (mainly self-employed).

Currently, industry codes are being assigned upon registration by the Corporate and Business Registration Department (CBRD) and the same codes are used by MRA. Statistics

Mauritius is working with these institutions in order to ensure that the codes are assigned correctly. This will facilitate the use of administrative source in the SBR. Description of activities should be obligatory when businesses register at CBRD. This will allow Statistics Mauritius to check and correct for codes.

Portugal

In 2006, a programme set up by the Portuguese government was in place – the Simplex Programme, whose main goal was to simplify and modernize the public administration. In that context, a new model was put in place through the joint efforts of four public entities that made it possible to acquire administrative data while simultaneously complying with four legal obligations.

The Simplified Business Information (IES) system is one in which related economic and financial data are collected from a single administrative source by electronic means and at a single point in time.

IES is considered a win-win project from which all the participants can benefit. For Statistics Portugal, its advantages include: a simplified data-collection process; a shorter time lag before data become available; access to more data with a complete coverage of the enterprise population; and a reduction in burden and costs – the structural business statistics survey was no longer needed.

For companies, this paper-free system makes possible a significant reduction in costs and response burden, because companies can fulfil four legal obligations once in a single process: the presentation of tax declarations to the Ministry of Finance; the settlement of accounts to the Public Register (Ministry of Justice); and the reporting of statistical data to both Statistics Portugal and the central bank. In 2015, a fifth public entity – the Ministry of Economic Affairs – joined this group, with the aim of gathering information on local units in order to update their commercial records.

As a result, the Portuguese Business Register has been receiving tax data for all companies (around 400,000) since 2006, making it possible to update the business sector (S11 and S12 – Non financial and financial corporations) fully on an annual basis.

South Africa

Business information contained in the SBR is updated from the South African Revenue Service, such as legal names, registration number, VAT and IT information, physical address, contact details, economic activities, and turnover.

Currently, the initial classification is obtained from South African Revenue Service, and further investigation is conducted by the SBR. There is a dedicated team focusing on the quality control of classification of all units investigated by components in the SBR. These components are responsible for the annual quality improvement survey; survey area feedback; VAT investigations and profiling.

The descriptions of economic activities are further requested in the different economic survey questionnaires.

Tunisia

The SBR was created in accordance with decree No. 94-780 of 4 April 1994, which entrusted the National Institute of Statistics of Tunisia (INS) with the creation of the register, and also its update, management and utilization and the dissemination of statistics.

It covers all patented enterprises (legal units) irrespective of the business sector of the enterprise, its size (employer or self-employed person) and its legal form.

The processes of creating and updating the SBR are mainly based on administrative files. The current main partners of INS in the management of the SBR are the Directorate General of Taxes (DGI) and the National Social Security Fund (CNSS).

6.3.2   Importance of business accounting data

6.36.   The enterprise's annual accounts are an important source of information about business activity and size. It is important for the national statistical office to have access to these data for the statistics production and for the SBR. The accounts can be used to learn about the activities, size, assets and also structural changes in the company, for example, regarding takeovers, mergers, split-offs and so forth. To this end, the national statistical office should have business accounting expertise within the SBR or profiling units for analysing and linking business accounting data. Gradually, as the accounts become digital, it is important that national statistical offices remain closely involved in the development of a data reporting solution and has legal and technical access so it can also be used statistically. As a minimum, the national statistical offices should work to have access to the digital data, and preferably as soon as they are available and reported to the administrative authority responsible for collecting and controlling company accounts (chamber of commerce, tax administration, business authority or similar). It may be daily, weekly or monthly, getting them annually is too rare.

6.3.3   Advantages of using administrative data

6.37.   The advantages of using administrative data to supplement or replace survey data wherever possible are well known. They are particularly pertinent to the construction and maintenance of an SBR, as outlined in the following paragraphs.

Coverage

6.38.   Use of administrative data in place of survey data eliminates sampling error, removes or significantly reduces non-response and provides more accurate and detailed estimates for various subpopulations, such as small geographical areas. This is because administrative sources often give complete, or almost complete, coverage of a target population, whereas sample surveys often cover only a relatively small proportion directly.

6.39.   Coverage is of great significance from an SBR viewpoint, given its aim of including all economically active units. Thus, the main advantage of using an administrative source is the level of coverage that it provides, which is either complete or, if not complete, at least well defined, so that it can be assessed against the target population.

Costs

6.40.   Another advantage that administrative data offer over survey data is their comparatively lower cost. Surveys are expensive, in particular if they are conducted as censuses or involve the use of personal interviews. Administrative data are often available free of charge, or for the marginal cost of extraction, in particular if they originate from the public sector. Even if there is a charge, it is often cheaper to use administrative data than collect the same information by survey. Fewer staff are usually needed to process the data and there is no need for non-response follow-up.

6.41.   The size and scope of an SBR make it very difficult and expensive to populate and maintain solely with statistical data. A periodic economic census would be required to build an SBR, which would be very expensive, as would be the complementary intercensal maintenance procedures.

Response burden

6.42.   Using data from administrative sources involves no additional response burden. While 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 – and therefore less necessary – burden. If they have already provided details to other government departments, they may be annoyed at receiving requests for similar information from the national statistical office.

6.43.   An associated advantage is that the use of administrative data may, in some cases, enable statistics to be produced more frequently, with no extra cost to businesses.

Content and timeliness

6.44.   The use of administrative sources may increase the quality of the SBR by providing access to more up-to-date information on key characteristics, such as:

- ▸   Name and address
- ▸   Births and deaths of units, and the dates of these events
- ▸   Economic activity code
- ▸   Location
- ▸   Size, in terms of number of employees or turnover or both

6.45.   In addition to improving the timeliness of SBR data in the context of their use for survey frames, administrative data can improve the timeliness of statistics that are derived from the SBR. This is because surveys take time to plan and design, to pilot questionnaires, to analyse the population and optimize the sample, to collect and process the data, and other processes. Access to a suitable administrative source through the SBR may provide a quicker and more efficient solution.

6.3.4   Disadvantages of using administrative data

6.46.   The following paragraphs outline some of the problems encountered in using administrative data for the SBR and methods that may be used to address them.

Administrative and statistical definitions differ

6.47.   Administrative data are collected for a specific administrative purpose, and the corresponding needs and priorities are likely to be different from those of the statistical system. Thus, administrative units may not coincide with statistical units, and their characteristics may be defined differently. For example, the tax authorities may permit a single enterprise to have multiple VAT accounts. Turnover collected for VAT purposes may not include turnover related to the sales of VAT-exempt goods and services, whereas the statistical system wants total turnover.

6.48.   Similarly, the classification systems used within administrative sources may be different from those used by the national statistical office. For instance, a register within a food safety authority might classify a department store according to its food department. Even if the classification systems are the same, they may be applied differently, depending on the primary purpose of the administrative source.

6.49.   Where classification systems are different, it is necessary to construct conversion matrices to map the codes in the administrative classification on to those required for the SBR. Such mappings may be one-to-one, one-to-many, or many-to-many. In the latter two cases, a probabilistic allocation may be used, but this is not desirable as

it does not give accurate codes for the individual units, codes which are needed for the SBR in its primary role of providing survey frames.

Timeliness

6.50.   Another common problem encountered when using data from administrative sources relates to timeliness. Administrative data may not be available in time to meet statistical needs, or may relate to a period which does not coincide with that required for statistical purposes. For example, the income tax year may not coincide with the calendar year required for structural business statistics.

6.51.   Furthermore, there is generally some time lag between the occurrence of an event in the real economic world and its recording by an administrative source. This is then followed by a further lag before the administrative data become available to the SBR. Lags in recording births and deaths of enterprises are a major source of SBR coverage errors. If these lags can be measured, allowance can be made for them in statistics based on SBR data.

Changes to administrative regulations or procedures

6.52.   Public sector administrative sources are set up for the purposes of administering regulations, for example collecting taxes or for 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 and other features. They may even be abolished completely. Such changes may happen quite suddenly, with little warning. Thus, reliance on administrative sources always carries a certain degree of risk in terms of continuity of supply.

6.53.   These risks can be mitigated by legal or contractual provisions, by obtaining early warning of impending changes through regular contact with those responsible for the sources, and by drawing up (and, where necessary, implementing) contingency plans.

6.3.5   Monitoring the quality of administrative data

Getting to know the administrative sources

6.54.   Although administrative data have many advantages, it is important to invest time in understanding and monitoring their quality. It can be useful to compare sources in terms of coverage of their units and accuracy of their characteristics. The closeness of administrative units and characteristics to the elements required for an SBR is an important factor in determining the quality of an administrative source. An SBR improvement survey may be required to determine the values of certain characteristics for which values provided by administrative sources are not considered appropriate.

6.55.   The starting point in assessing the quality of an administrative source is to build up a thorough knowledge of the source, including its primary purpose and the way in which the data are collected and processed. Thorough understanding of a source enables a more accurate assessment of its strengths and weaknesses. To help develop and document this knowledge, in particular for the benefit of future SBR staff, it is useful to develop some form of template to record information from the source on contacts, units, characteristics, quality and formats. As noted in section 6.2.1, this information should be discussed with the administrative authorities and should be reviewed regularly.

Quality indicators

6.56.   Regular monitoring of quality is important. A full range of quality indicators should be developed. Examples of quality indicators include the following:

> ‣   Number and proportion of enterprises lacking a valid and complete economic activity code.

> ‣   Number and proportion of enterprises for which the activity status (active, dormant, dead and so forth) is unknown.

> ‣   Number and proportion of enterprises lacking a complete address.

6.57.   More examples of good practice are provided in annex D.

6.58.   Such indicators may be compiled using feedback from surveys based on the SBR or from SBR improvement surveys.

Dealing with conflicts

6.59.   Where an administrative source and the SBR do not agree, the reasons should be investigated with the aim of gaining a better understanding of SBR quality. Surveys may be used to investigate such discrepancies. They may be conducted, either specifically for this purpose or as part of some other data-collection exercise. The investigations should help in determining the appropriate rules and priorities for updating from different sources.

6.60.   There may also be conflict in data from different administrative sources, for example conflicting industry codes. Procedures and rules need to be developed to resolve these problems. This could involve verifying the data by contacting the enterprise or undertaking analytical work to determine which source is most reliable. The goal is a set of general rules to deal with conflicts.

6.3.6   Legal issues and access to administrative sources

Access guaranteed through statistics act

6.61.   The first step in the use of administrative data is to ensure that the national statistical office has access to the data. There are two aspects to gaining access: the first is the legal framework, the second is setting up and implementing the procedures for transfer of the data. While the legal framework for statistics varies from country to country, the preferred approach is for the national statistical office's right of access to administrative data to be enshrined in a general statistics act. Transfer of data is then simply a technical issue, with strict rules on security to be followed.

6.62.   In many countries, the national statistical office does not have legislated access to data. There may even be confidentiality legislation preventing the transfer of administrative data to the national statistical office, even for statistical purposes. This can prove to be a major problem in establishing a good SBR, and one that may not be solved quickly. While the best approach is likely to vary according to the country, in any case the national statistical office should consider initiating and introducing changes to its statistical legislation to ensure access is guaranteed by law. Box 6.5 presents some examples of national statistical offices successfully gaining access to administrative data.

Formal agreements with administrative data suppliers

6.63.   As previously noted in section 6.2.1, whether or not the national statistical office has legislated access, the national statistical office should try to establish some form of formal agreement with administrative data providers. This can be a legally binding contract with a private sector supplier, or a service-level agreement or a memorandum of understanding within a public sector provider. Such agreements should describe the rights and responsibilities of both parties, delivery flows, data confidentiality constraints, quality standards, frequency and format of data transfer, time frames for responding to queries and questions about the data, and procedures to follow in case of disputes.

Box 6.5
Examples of national statistical offices gaining access to administrative data: Brazil, Colombia, Croatia, Mexico, Mongolia and Tunisia

Brazil

IBGE, despite coordinating the national statistical system, does not have access to tax data, which are protected by fiscal confidentiality. Thus, the number of employees is adopted as the stratification variable in the sample selection of the economic surveys, despite the limitations of this metric.

Where formal activity is concerned, the number of employees in the local units of each legal unit in the national territory is obtained through an agreement with the recently dissolved Ministry of Labour, which allows access to the Annual Report of Social Information (RAIS), an annual administrative record which contains data on employment, with a particular focus on the demographic characteristics of the workers.

In the Central Business Register (CEMPRE), the Brazilian SBR, some 95 per cent of the registers are updated through RAIS, which supplies information on establishments at national level, but the annual periodicity imposes a lag of at least eight months on CEMPRE information when the surveys are being collected.

Two strategies were adopted to mitigate the effects of this discrepancy. With regard to the number of employees, the agreement with the Ministry of Labour was extended to cover the National Registry of Employees and Unemployed (CAGED), a monthly administrative register that logs occurrences of hiring and dismissal in the country. It is used to update the information on employees, reducing considerably the gap at the moment of sample selection. Following to the recent relaxation of labour laws, however, which now allow formal part-time temporary hiring, the annual total of the monthly balance of hiring and dismissal occurrences differs significantly from the number of employees recorded in RAIS. Thus, in order to ensure the consistency of CEMPRE, CAGED is only used to update the survey frame, not the SBR.

Where non-economic information is concerned (enterprise name, address, contact information, etc.), a comprehensive project is under way to obtain alternative data sources. Automatic web-scraping routines and periodic API consultations are being developed, along with the establishment of new agreements with institutions holding registers of specific populations, an initiative facilitated by the publication, in 2016, of a decree establishing rules favourable to data-sharing between entities of the public administration. This decree, however, excludes tax data.

Colombia

For the updating and maintenance of the SBR, cooperation agreements are concluded among the entities concerned, in order to provide information on the totality of units that carry out activities in various sectors of the economy. These agreements specify the periodicity, type and validity of the information, to ensure that it is comparable across time for the statistics of the economy.

Box 6.5 (*continued*)

In addition, since 2015 information has been exchanged between the information-producing entities belonging to the National Statistical System (SEN) through Decree 1743 of 2016, which facilitates access to administrative records for statistical purposes.

Currently, to update the SBR, agreements are maintained for two years with entities that are not part of the National Statistical System and with entities that require a formal agreement before any data can be shared, as is the case of the Ministry of Health. With the other entities, the exchange of information does not expire. In both types of agreements, the definition of the variables must be explicitly declared.

Croatia

The situation regarding administrative sources in Croatia is very favourable for the needs of the SBR. Several administrative bodies provide data on registration and the main characteristics of units, while other sources provide data on turnover, employment and links between units (ownership and control).

The Administrative Business Register has existed since 1976 and is maintained by the Croatian Bureau of Statistics. It includes all legal persons that operate in Croatia and represents an exhaustive source for trading companies, public and government units, non-profit organizations and all other legal forms.

Craft businesses constitute another kind of unit, which represent a significant part of the Croatian economy (one third of active units).[42] These are classified as the economic activity of natural persons – the owners of the craft businesses – and are registered in the Crafts Register.

Those two administrative sources cover some 95 per cent of active legal units and provide data on the main identification and contact characteristics, and also the activity code. The remainder of the units are natural persons – independent professionals who are registered in a number of different registration bodies – and, for that reason, the Register of Tax Payers has been chosen as a unique administrative source for units of this kind.

Economic details are provided to the SBR by the Register of Annual Financial Accounts (maintained by the Financial Agency) and the Tax Administration (VAT, income and profit tax, unique form for tax, surtax and other salary-based contributions). The Commercial Court Register and Central Depository and Clearing Company Inc. are the sources of information on the ownership of units. Several other sources are used for the identification of local units.

Good cooperation and agreements with all the above-mentioned bodies and organizations are essential for the timely delivery of the complete set of the necessary data.

The Croatian Bureau of Statistics concludes memorandums of agreement with administrative sources, defining the content, structure, format, periodicity and deadlines for data delivery.

The most important factor that enables the use of administrative data is the existence of a common identifier used by all institutions. In Croatia, there are two main identifiers: OIB (personal ID number –for both legal and natural persons) and ID number, which is issued by the Administrative Business Register and makes it possible to consolidate data in the Statistical Business Register.

Once the data are received in the SBR, they are checked for mistakes by comparing data from different sources (such as annual accounts and taxation data) and by comparing consecutive years.

If potential mistakes are detected (usually not many), they are reported to the administrative source for checking and correction.

[42] This is a specific type of businesses in the countries of the former Yugoslavia. Most crafts businesses are small businesses owned by natural persons. In Croatia, the majority of them undertake market-oriented activities in industries such as retail trade, catering, personal services, construction and so forth.

Box 6.5 (*continued*)

The economic activity code is initially assigned to legal units by administrative sources, but it is sometimes changed in the SBR based on the information received from statistical surveys of business statistics. Information of this kind is considered to be statistical and confidential and it is not reported back to administrative sources for correction. Nevertheless, the Croatian Bureau of Statistics actively provides methodological support to administrative registers in applying the national activity classification in their registration processes. In that way, the quality of activity codes has significantly improved.

Nowadays, almost all administrative data are received by web services in the agreed structure, which enables efficient processing. The good quality of data is not only a result of the interconnection between administrative sources and information-sharing, but also of the feedback provided by the Croatian Bureau of Statistics.

Mexico

Among the several administrative sources used to update RENEM, two administrative registers in particular are provided by the Ministry of Economic Affairs and the Ministry of Tourism. Agreements have been concluded with both agencies to allow INEGI to host in its servers the applications and databases of the Mexican Business Information System (SIEM) and the National Tourism Register (RNT).

In both cases, the first task was to review their register´s conceptual and informatics structures, and also the requirements derived from their administrative functions established in their respective legislation, standards and regulations. This has enabled INEGI to standardize the fields of each of the common variables with RENEM and incorporate the statistical variables that it requires.

The linkage of RENEM with these registers starts with automated matching and identification tasks using standardized variables, such as the location and name of the natural or legal persons and the industry activity.

Subsequently, a more detailed manual revision is carried out. In the particular case of RNT, the observation unit and the specific industry activity are carefully reviewed to guarantee their correct linkage with RENEM, to avert overlapping economic units and misclassification in their North American Industry Classification System (NAICS) code. This register considers as independent the different services provided by the same establishment, in contrast with RENEM, which considers just one activity. This means that one establishment in RENEM can have two or more tourist services providers which are linked together: for example, a hotel with integrated services can be related to a restaurant, spa, shop and travel agency, all of them certified independently by RNT.

The linking of both registers enables RENEM to be continuously and directly updated with the business data of the specific sectors, reducing the burden on the respondents. Furthermore, INEGI is exploring the possibility of establishing more institutional arrangements with other institutions to improve and update the SBR.

This will also form an important basis for compliance with the goals set by the National System of Statistical and Geographical Information (SNIEG) on the generation of statistics through the application of administrative records.

Mongolia

Since 2018, the national statistical office of Mongolia has organized a series of activities such as meetings and presentations aimed at integrating the SBR database with administrative databases and advocating the importance of integrating the SBR database to the owners of the administrative databases. As a result of these activities, several agreements have been signed with various government agencies. Furthermore, the legal environment for connecting the legal entity register, owned by the General Authority for State Registration of Mongolia, has enabled the national statistical office of Mongolia to develop and start to use the software tools for the integration of these databases into the SBR.

Box 6.5 (*continued*)

There are many advantages in integrating the SBR with the legal entity registration and other databases of government agencies. For instance, when an entity registers for the first time as a legal entity in the legal entity register, this registration will appear automatically in the SBR at the same time, which improves the coverage of the SBR. In addition, the entities do not need to report their legal type, form of ownership, details of their directors and other information that is already available to the SBR in the legal entity register, thus reducing the response burden. Furthermore, the national statistical office of Mongolia organizes training sessions for the registrants to improve the quality of administrative data by promoting the inclusion of correct codes and definitions in the SBR and legal entity register databases, such as the legal type, form of ownership and type of activity of an entity.

Tunisia

Several administrative sources containing business databases can be used for setting up and updating the SBR. The main problem, however, is that of coverage, whether it is exhaustive or not.

In Tunisia, the most exhaustive source for companies is the Directorate General of Taxes (DGI). It contains all the legal units (licensed companies), because all companies must be entered in DGI in order to carry out their business activity.

Other administrative sources have coverage problems; in fact, these sources all cover a part of the DGI file, for example:

> The Tunisian National Social Security Fund (CNSS): this is the most extensive source after the DGI and it covers most of the employers and a good part of the self-employed.

> Customs: this only covers companies that have carried out an import or export operation.

> Agency for the Promotion of Industry and Innovation: this only covers companies that have an interest in registering with the Agency (usually to benefit from State benefits).

> The commercial register: this only covers companies carrying on a commercial activity.

Although the integration of several sources in the SBR is useful for enriching content and improving quality, the most essential source in Tunisia is DGI, which constitutes the most exhaustive source of information on companies. In Tunisia, there is an annual update of the SBR and the administrative data must therefore be obtained in a regular and continuous manner. To this end, agreements with the partner administrations have been drawn up (primarily DGI) based on the statistical law and the decree establishing SBR, which grant INS access to the various administrative sources containing useful information.

6.64.   Even if the situation differs from one country to another, there is usually more than one administrative source available. Figure 6.1 lists some of the main criteria that might be applied for the actual selection of the administrative sources. Depending on the extent to which these criteria are met, it should be obvious which source or which sources are important for the maintenance of the SBR and should be accessed. Of course, not all sources might be used from the beginning onwards; on the contrary, it is advisable to start with the most relevant source and add further sources as required in a stepwise procedure. The criteria listed in figure 6.1 should be applied in the national context.

Figure 6.1
Criteria for the selection of administrative sources

Criteria	Explanation
Coverage	A source that covers a broad range of the intended SBR population is preferable to a source that only covers a smaller part of the population, especially when starting with the use of administrative data. In that sense, it is clear that, for instance, tax registers are an overall important source.
Relevant variables	Administrative sources usually have basic identification data on the units in the population. These relate invariably to the name of the unit, the location and address and probably also the legal form and information on the economic activity (based on the international statistical classification or on a national classification). Further data on the units are needed for the SBR, however, such as the number of employees and the turnover, as two of the most important stratification variables. Usually tax data will provide turnover figures, but not necessarily employment data.
Register unit and identifiers	It is not be expected that, in administrative sources, the register unit will coincide with the statistical units needed in the SBR. If, however, the administrative source is based on the legal unit for most of the units, this could directly be used as the enterprise unit. This might not apply for big and complex enterprises. The same holds true for other units in the administrative register, such as the local unit; if their definition is identical or close to the statistical definition.
Frequency	The updating of the SBR should, as far as possible, be a continuous process. A data source that is available on a monthly or quarterly basis is thus better than sources that are only available annually.
Timeliness	Not only the frequency is of importance but also the timeliness of the register information. If, for instance, administrative data can be accessible on a monthly basis, but the updating of these data is done less frequently, then the higher frequency of access is of no real use. Furthermore, it is highly relevant whether the updating process covers all register units regardless whether an actual change has occurred or not, or whether the updating takes place only on a case-by-case basis.
Quality	Quality of the administrative data is a further criterion that needs consideration. The quality problems may be given for various reasons and may affect all or just certain specific register units or just certain variables. Since the coverage and relevant variables vary over time and are even susceptible to political changes, the quality of the administrative data is not always stable. As there is normally no data source without some quality problems, the quality needs to be continuously evaluated (see annex D3 for an evaluation checklist). Quality problems should not per se be a reason to exclude a specific administrative data source. On the contrary, together with the owner of the administrative source, procedures should be developed as to how the quality might be improved.

Building relationships with administrative data suppliers

6.65.   In addition to formal arrangements, good working relationships with administrative data providers should be developed. These can be achieved through regular contact, preferably face-to-face. It is usually worth devoting time to visiting suppliers to gain greater knowledge of their work, systems and constraints. This will lead 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.

6.66.   In some countries, administrative agencies may be willing to release data to the national statistical office. In other countries, there may be reluctance to do so. It is very difficult to deal with such concerns, but possible approaches include the publication of clear limits and rules regarding the national statistical office's use of data, thus ensuring that businesses understand that sensitive data will not be fed back to other parts of government (in particular, tax agencies).

6.67.   It is advisable that the national statistical office should be coordinated when it approaches administrative sources: it should not be the case that different departments of the national statistical office approach the same source in order to receive more or less the same data. There should be coordination within the national statistical office. In this context the SBR, as the primarily user of such data, has a specific role.

6.68.   A favourable situation is seen in countries, where the statistical legislation not only grants access to administrative data records in general, but also stipulates that public sector authorities are required ex ante to involve the national statistical office in the preparation and establishment of new registers or other data sources also usable for official statistics and the amendment of existing registers and systems. In some countries, such as France and Denmark, the SBR is established and maintained according to a separate national legal act stipulating the data content and, not least, the responsibilities for maintaining the data from one or more administrative sources for the SBR, and the usage of the data for statistical and non-statistical purposes (see also chapter 9, section 9.3.2). The optimal situation is when the definitions of units and variables, the classifications and other elements used in the joint public registers comply with the international statistical definitions and recommendations.

Box 6.6
Legal framework concerning access to administrative data

As explained above, an adequate legal framework based on the Fundamental Principles of Official Statistics is an indispensable requirement for developing and maintaining an SBR and the production of high-quality statistics. In order to support countries in developing adequate legal frameworks for their official statistics, ECE has developed recommendations and good guidance (see the Generic Law on Official Statistics for Eastern Europe, Caucasus and Central Asia, published in December 2016, available at https://unece.org/statistics/publications/generic-law-official-statistics).

One of the pillars of such a framework is legal access to administrative data. Article 17 – "Access to administrative data" – of the Generic Law provides a model text for the national statistics law:

> 17.1   All national and local authorities are obliged to provide the *Producers of Official Statistics*, free of charge, with data in their possession at the level of detail necessary for the production of official statistics and with metadata that enable assessing data quality. Special confidentiality or secrecy provisions in other legislation cannot be invoked unless the legislation explicitly excludes the use of data for statistical purposes.

> 17.2   If the providers of administrative data plan to develop a new data collection or carry out a major revision in their data collection or processing in a way than may affect data provided for official statistics, they shall consult the *National Statistical Office* and, where appropriate, the *Other Producers of Official Statistics*, in advance of the decision.

The inclusion of such a prescription in the national statistics law, as outlined in the above article, will reinforce the legal position of producers of official statistics to receive administrative data. The providers of administrative data have to provide the data needed for statistical production free of charge at the necessary level of detail and with the relevant metadata. The individual data flow follows a one-way principle, whereby individual data may not, for any reason, be provided from the producers of official statistics to administrative data providers. The producers of official statistics may, however, cooperate with the administrative data providers to improve their procedures and methods of quality control and error corrections. In addition, providers of administrative data have an obligation to consult the national statistical office, and any other producer of official statistics affected, in advance and as early as possible about any expected changes in their data collection or processing that may affect data provided for official statistics.

6.4   Using administrative sources in practice

6.69.   This section provides guidelines on the steps to be taken when creating and maintaining an SBR primarily using administrative data. The precise sources used depend on what sources are available and on data access. The following basic principles apply whatever the sources.

6.4.1   Keeping administrative and statistical data separate

6.70.   Data about administrative units received from administrative systems should be stored separately and not mixed with data for statistical units.

6.71.   Even in the case of combined administrative and statistical registers (see sect. 6.2.4), it is important to maintain clear procedures on data sources and updates to ensure that data received from administrative records and any statistical transformations are auditable.

6.72.   Experience shows that it is much easier to discuss unusual or anomalous SBR data either with the enterprises themselves (which should know what information has been given to the administrative sources) or with the administrative sources directly, rather than to try to solve the problems within the SBR. Keeping information separate allows specific problems to be discussed and specific examples to be provided if investigation is required. The particular sources of data that should be used in the SBR, both administrative and statistical, depend upon the outcome of such investigations and the feedback received from the relevant sources.

6.73.   Another possible reason for separating administrative and statistical data is if administrative data are being used for other statistical purposes. For example, social security data may be used not only for SBR purposes, but also to provide employment data for a variety of statistics.

6.4.2   Establishing unique identifiers

6.74.   The establishment of unique identifiers (as further discussed in section 11.4.10) is essential for the accurate maintenance of the SBR. Use of administrative numbers for statistical units may lead to added complexity in SBR systems, and introduce risks such as duplication and omission of statistical units. Thus, an identification numbering system should be created covering each type of statistical unit.

6.75.   Where multiple administrative sources are used, there may be several different identifiers, one for each source. In this case (as further discussed in section 6.8), significant resources may be required to match and create unique enterprises using a combination of data from these sources. If the national statistical office has influence over the development of administrative data, it should promote use of a unique business number for all relevant administrative processes in the country. While this may not be easily achievable in practice, there are countries where it has happened.

6.76.   Administrative units – if they are not already in a one-to-one correspondence with legal units – should first be linked to the corresponding legal units. The most complex situation is one where a legal unit may have several administrative units within the same administrative source and hence have several identification numbers in that source. Furthermore, the legal unit may close one or more of these administrative units or ask for more to be created. It is important to handle these situations carefully. Usually, it is possible to identify the main unit in the administrative register as the one to be used to provide information about the legal unit.

6.77.   It is important to see the difference between legal units that are economic producers and legal units that are non-productive but may own other legal units in an inter-corporate ownership structure. By law, a legal unit is not legally responsible for another legal unit even though it may be part of the same enterprise group and control the other legal unit. It is therefore vitally important to identify all legal units within an enterprise group by all necessary means, even though some of them may not be readily identified from administrative sources.

Box 6.7

Creation of unique IDs in the administrative sources in Mauritius

The Business Facilitation Act of Mauritius provides a framework that allows any business to start operations on the basis of self-adherence to comprehensive and clear guidelines. The Corporate and Business Registration Department (CBRD) is responsible for incorporating and regulating companies (including offshore companies) and all persons doing businesses in Mauritius. "Person" is defined as including a consortium, company ("société" in French), partnership, joint venture, trust or firm.

The Business Registration Act, which came into force in October 2006, has provided an important tool for coordination, link, control and harmonization of economic activities. The business registration number (BRN) which is assigned to operators at the time of registration uniquely identifies the operator and is intended for use by organizations such as local authorities responsible for issuing operating licences, the Mauritius Revenue Authority (dealing with customs, income tax, VAT), the Ministry of Social Security and others. From 2006, any business operator is required, by law, to apply to CBRD for a BRN before requesting a licence to operate (from a local authority) or registration with a regulator (from a regulatory body).

BRNs are issued at enterprise level and not at establishment or activity level. Thus an enterprise comprising more than one establishment will be allotted only one BRN, irrespective of the locations of the different establishments. For instance, a bank which has several branches across the country will have the same BRN for all the branches. Likewise, an establishment involved in more than one type of economic activity at the same location will have only one BRN.

Box 6.8

Common administrative business register in Denmark

One way to get better data in the SBR is to have fewer, but more authoritative and robust, data sources. It is often an efficient strategy to bring the stakeholders together and identify the common needs of those public authorities – and private sector users – who are the main users of basic business data. Everybody is interested in high quality and this is exactly what a common business register with agreed roles and responsibilities for updating can ensure. Most of the stakeholders also have some special requirements, which they then have to manage in their own part of the register system. In Denmark, for example, a common basic register was created that enables companies to register '"only time in one place", so that information about the company's lifespan (birth, death), main activity, address and addresses for local units, legal form and, in particular, a common ID number can be linked to one's own number series. The structure of such set-up is shown in figure 6.2.

Figure 6.2
Danish administrative business register set-up
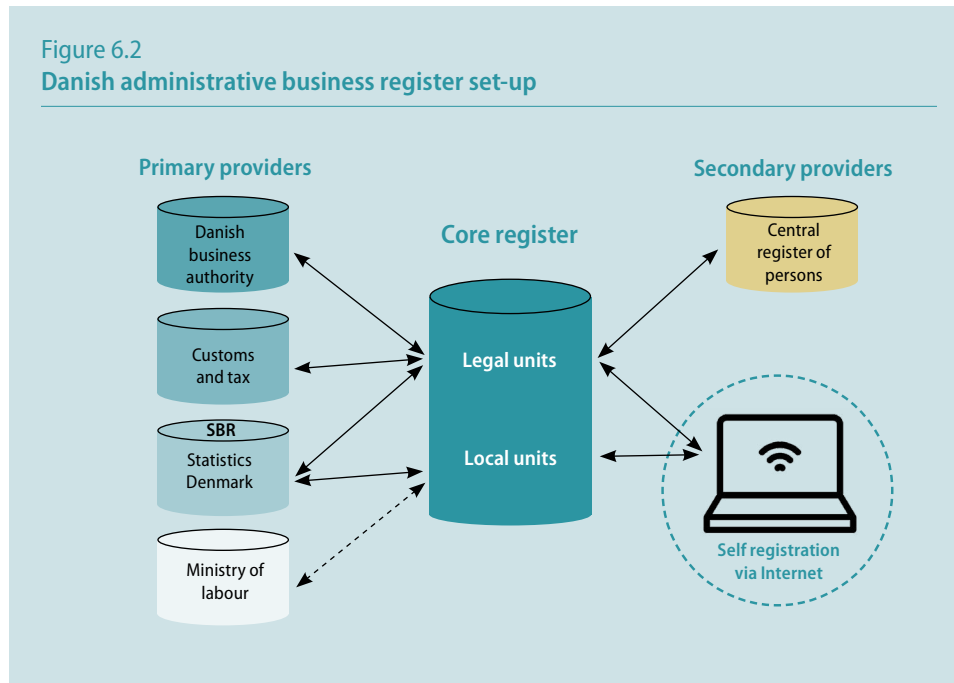
6.4.3   Using sector-specific and additional sources

6.78.   As previously noted, some trade associations, charity regulators and government ministries such as those responsible for education and health may maintain data that are useful in providing coverage of certain sectors of the economy and additional content. Care must be taken, however, to ensure that coverage of the particular sector is complete or, if not, that the gaps are understood and that the possibilities of duplication and omission are carefully considered. In addition, using multiple sources leads to challenges in maintenance and management. In some cases a source may be used to verify data from another source; in other cases a source may contribute directly to updating of statistical units.

6.79.   Examples include the following:

> ▸   The characteristic sales space may be available from an industry association for retail trade enterprises.

> ▸   The characteristics "category and number of stars" and "number of beds" for hotels may be available from a tourism authority or bureau.

> ▸   Import and export data may be available from the customs authority.

6.80.   Such administrative sources may be used in data confrontation and also in validating statistical units. They can increase sampling efficiency (by reducing the number of units sampled for the same level of accuracy of survey results) in two ways:

> ▸   The SBR includes links to units in these sources that can thus be consulted and that contain additional stratification characteristics.

> ▸   SBR data are of improved quality thanks to cross-checking with these sources.

6.81.   Use of additional sources can also reduce the response burden on enterprises through increased sampling efficiency or by removing the need to conduct a survey altogether. The sources may even include additional units not present in the basic administrative sources, thus extending the coverage of the population of interest. At

the same time, the priority rules for determining the values of those characteristics which are available from more than one source need to be determined.

6.82.   To ensure that a source is sufficiently well aligned with the SBR to be easily usable, it may be necessary to introduce additional criteria for its use, such as common unit identifiers, common definitions and classifications. The greater the coherence, the more useful the source is likely to be.

6.83.   The rights and mechanics of access to administrative sources are also factors in facilitating their use. Many potentially useful sources are likely to exist, but they are of no actual use if the legal and technical means to link them to the SBR are not in place. Thus, where the data are deemed to be of considerable potential use, the national statistical office should invest resources to enable the appropriate links to the SBR to be made.

6.4.4   Thresholds

6.84.   It is important for the SBR to have as complete coverage as possible, so care needs to be taken to understand the impact of thresholds. For example, an enterprise may not be required to register for VAT if its turnover is below a certain threshold. If the threshold is low then the enterprises that are omitted because they are below the threshold have only a small impact on the overall estimates and their omission from the SBR is not an issue. If, however, the threshold is quite high, as it is in some countries, then its impact on coverage and estimates is likely to be large. In this case, additional sources and methods may be needed to supplement coverage provided by the source. This typically requires matching and linking the additional sources to the SBR to avoid duplication and minimize omissions.

6.4.5   Handling changes in administrative sources

6.85.   Changes in administrative sources, for example, changes in thresholds, definitions or computer systems, can lead to discontinuity in the data supply. Critical to managing this type of change is working very closely with the administrative source to gain a full understanding of the change in order to assess the impact on the SBR. If possible, it is good practice to measure the impact of the change, for example by processing the administrative data following the change in a copy of the SBR database so that the impact can be easily identified before regular processing takes place. Often there are unexpected consequences of changes to administrative data, in particular when complex systems are involved.

6.86.   Once an assessment of the impact has been made, the next step is to work actively with the survey staff, who are the most important SBR users, to assess the impact on their outputs.

6.87.   Changes in tax thresholds are common and often just reflect the general inflationary environment in the country. This type of change may not have any real impact on coverage. Occasionally more radical changes in thresholds take place, reflecting a political decision, and the change may well be permanent. These types of changes have an impact on SBR coverage, particularly if a single source is used. It may be possible to estimate the impact on the population, for example by comparing unit birth trends before and after the change.

6.88.   There are two possible methods of addressing changes so that they do not distort the statistics provided to users:

> ‣ Make an appropriate transformation in the SBR to minimize the discontinuity.
>
> ‣ Let the survey staff deal with the discontinuity through their own systems.

6.89. The method used must be clearly linked to the nature of the change. For example, a change in the classification systems in administrative data can best be handled by transforming data in the SBR, whereas an expansion of the SBR frame population caused by improved SBR coverage procedures may have to be handled by the survey staff.

Box 6.9
Country examples on dealing with multiple sources: Colombia, Mauritius and the State of Palestine

Colombia

The National Administrative Department of Statistics has prepared a guide to documenting the methodology of statistical operations based on administrative registers, which provides guidelines for the documentation of statistical operations based on an administrative register with standardized and harmonized criteria for their structure, presentation and content.

In addition, the Department has created a guide for the elaboration of metadata of the administrative register, providing guidelines for the construction of the metadata for registers to promote their use, comparability and integration. See www.dane.gov.co/files/sen/lineamientos/DSO_020_LIN_04.pdf.

Mauritius

All fields in the SBR of Mauritius, except for "identification" and "remarks", are associated with a list of sources which can modify them. The list is in order of priority, the first source in the list having highest priority. If a field already has a value, it can only be changed by a source having a higher priority than the source of the existing value or a source of equal priority but with more recent data.

For example, if a field has an associated source list comprising the sources: "VAT", "Companies division" and "Social security", in that order, and the field already has a value from "Companies division", it can later be modified by data received from "VAT" or more recent data received from "Companies division". The system will, however. ignore updates from "Social security" or older data from "Companies division" for that field.

State of Palestine

PCBS recognizes that there is a difference in the quality of a certain variable among different sources, and there is also a difference in the quality of different variables of the same source. Thus, a list of ranking priories should be implemented to indicate the source of each variable in the SBR (such as commercial name, address or activity) and from where it should be taken.

The following example shows how the priority ranking of the sources is used to choose the variables: the ID and the commercial name of companies are taken from the Ministry of Economic Affairs, the owner name of individual establishments is taken from VAT as the first priority then from the municipality, while the address data are taken from the census, or the municipality (second rank), or the VAT file (third rank).

6.5   Identifying statistical units

6.5.1   Identifying enterprises

6.90.   In most countries with administrative registers, a legal unit of some sort is defined for SBR purposes. In many situations the legal unit is exactly, or a close approximation to, an enterprise. However, it is still necessary, however – at least for the bigger units, especially for those forming part of an enterprise group – to be able to create an enterprise in accordance with the enterprise definition, which does not require it to be in one-to-one correspondence with a legal unit.

6.91.   Whatever the particular definition or profiling approach, the transformation from legal units to statistical units must follow well-defined basic rules. The rules must make it clear how the administrative units from each source contribute to the updating of the statistical units, which source has data that may overrule or overwrite data from another source, and when the changes are deemed to have happened.

6.5.2   Identifying local units and/or establishments

6.92.   Although in principle there is nothing to stop an SBR containing both local units and establishments, typically, for resource reasons, it contains one or the other. Local units and establishments enable regional economic analysis to be conducted.

6.93.   Many of the usual administrative sources, such as VAT, refer to a legal unit as a whole or a specific part of it, but not necessarily at local unit or establishment level, so they cannot be used to populate local units and establishments in the SBR. In a few cases, local units are directly registered in administrative sources. Establishments never are. For that reason, creating local units and establishments is a bigger problem than creating enterprises.

6.94.   Some administrative sources are more useful than others. For example, employment tax systems, or social security databases, may hold information on the location of the employees – this information could be used to construct a local unit. Other sources may list business sites for business tax purposes. Whatever sources are used they have to be linked to or matched with other data sources or the SBR or both.

6.95.   As noted above, some administrative sources identify local units. If this is the situation, then it is important to establish whether these units are defined in the same way for all units covered by the administrative source and, hence, whether the data can be used as the basis for creating local units in the SBR.

6.96.   In most countries, there may be no suitable administrative data source and local units and establishments have to be established through a special SBR survey. As many small enterprises never have more than one site, it is best to stratify such a survey to ensure that all larger businesses are included, whereas smaller businesses are sampled very lightly, simply to estimate the probability of smaller businesses owning more than one site. The resulting data can be used to prepare a typical model for small businesses.

6.97.   Even so, a survey is an expensive option and very burdensome for the enterprises, since a range of characteristics have to be collected for each local unit or establishment. To reduce the associated burden and expense it may be possible to obtain the same information through an annual employment survey.

6.98.   Countries without local unit or establishment data should analyse the cost-benefit trade-off, in order to make an informed decision about identifying these units. Among the benefits are more granular geographical data, businesses that are industry-coded at this most detailed geographical level, the potential to geocode establishments to generate maps (for example of business location or employment intensity), and an

SBR that can serve as both an establishment and a more aggregate economic-unit sampling frame. The costs to be considered include both the direct costs of obtaining this information through a survey or a periodic census, and the cost to businesses in response burden

6.5.3   Converting legal units to statistical units

6.99.   The following paragraphs consider the processes that are necessary to transform data from administrative sources into the creation and updating of statistical units and their characteristics in the SBR. The main processes are listed and explained briefly. They need to be adapted to fit the specific requirements of each country and source.

6.100.   A principle adopted by several national statistical offices is that the SBR should serve as the gateway to administrative microdata. This means that all administrative data at the level of individual enterprises (or local units) should be fed through, but not necessarily stored in, the SBR. This enables the units and characteristics in each administrative data source to be matched to those in statistical surveys based on the SBR.

6.101.   Another approach is to use the administrative units as observation units and, through the links in the SBR, to transform the administrative data into statistical data at the level of each individual enterprise or establishment.

6.102.   The first step in processing generally involves checking the quality and coverage of the incoming administrative data to ensure that some basic conditions are satisfied, for example that:

  ‣   The file is the expected size; in other words, it contains roughly the expected number of records and the required characteristics.

  ‣   The values of the characteristics are in valid formats and ranges or both, for example, dates are within a permissible range, text fields contain only text characters, numerical fields contain only numbers, and codes used are valid.

  ‣   There is good coverage of the main characteristics, for example, identity numbers, addresses and economic activity codes are present for all units.

6.103.   If the preliminary analyses show that the incoming data are of sufficient quality, the next step is to transform them into updates to statistical units.

6.5.4   Transformation rules

6.104.   An administrative source is unlikely to use exactly the same definitions of units and characteristics as the SBR, especially in the case of complex businesses. The transformation of administrative unit data into statistical unit data may therefore involve several steps, including, first, the creation of the corresponding statistical units (typically enterprise groups, enterprises and establishments), followed by the derivation of their characteristics. Most enterprises have a one-to-one relationship with the corresponding administrative units. For large, complex enterprises, the relationships are likely to be more complicated and based on the results of profiling by means of which related legal, administrative and operational units are identified and appropriate statistical units are created. Derivation of the characteristics of statistical units usually involves the creation and use of algorithms or look-up tables to convert values of classifications and other characteristics obtained from administrative units to those appropriate for statistical units. An example is the conversion of economic activity codes that are peculiar to a particular administrative source to the standard codes used in the SBR (ideally, ISIC Rev. 4, or equivalent). The resulting transformation rules are dependent on the circumstances in each country.

6.105.   As already mentioned, it is important to determine whether it is necessary to put in place an intermediate step in the transformation from administrative to statistical data, by creating legal units to which both administrative units and statistical units are linked. Even if this is done, it might be difficult to automate the transformation from legal unit to enterprise as, in certain cases, there may not be a one-to-one relationship.

6.5.5   Automated updates

6.106.   In an ideal situation, the SBR would receive regularly scheduled and automated updates from administrative systems. In this situation, validation processes would need to be built into the programmes, using the same rules for creating and updating statistical units. Methods for identifying anomalies are dependent upon specific country processes and requirements, and on SBR resources.

Box 6.10
Main results of data exchange and data sharing in ECE couintries

A statistics law can be both an enabler and an obstacle to data exchange among producers. Better possibilities of secure data exchange, nationally and internationally, could help some statistical offices to enhance the quality, coherence and relevance of economic statistics and the efficiency of their production. Part of this exchange can be carried out at the aggregated level, but it will be necessary to find solutions enabling individual data to be exchanged in a secure environment for statistical purposes only. Both legal and technological solutions will need to be sought to enable such data exchange in a highly controlled environment, while avoiding any risks to privacy and fostering the trust of respondents and the public.

In 2016, Statistics Finland and ECE carried out a survey of countries to review the current practices of data sharing for statistical purposes, at national and international levels. In total, 48 statistical offices replied to the survey. The exchange of data on multinational enterprise groups (MNEs) is still relatively rare. Every fourth responding office had examined the activities of MNEs with other countries and every third office within a country with other producers of official statistics in their country.

National laws regulating data-sharing exist in 90 per cent of countries that responded to the survey. A common business identifier is used in over 75 per cent of countries. The protection of confidential data is well ensured in the legal frameworks. Sometimes data exchange is agreed upon and defined in the statistical work programmes. Agreements on the provision of administrative data to the producers of official statistics are very common with various administrative data providers. It is becoming more common that the statistics law provides a mandate for access or an obligation to use administrative data sources for statistical purposes.

One way to address the legal obstacles associated with data exchange is to help countries to draft legislation that facilitates data exchange. The ECE Task Force on common elements of statistical legislation (co-chaired by Latvia and the United Kingdom) drafted the ECE Guidance on Modernizing Statistical Legislation[43] for all statistical offices and it also touched upon data exchange. The guidance also reflects views presented by the Task Force on exchange and sharing of economic data at the time of drafting the recommendations related to data exchange.

While in some countries statistical legislation may not allow the exchange of individual data even among the producers of official statistics, the guidance recommends sharing of data for statistical purposes. First of all, the guidance recommends a mandate for

---

[43] United Nations Economic Commission for Europe, Guidance on Modernizing Statistical Legislation (Geneva, 2018). Available at https://unece. org/statistics/publications/ guidance-modernizing-statistical-legislation.

Box 6.10 (*continued*)

exchanging individual data among the entities belonging to the national statistical office exclusively for statistical purposes in the respective area of competence of each producer of official statistics.

The guidance also supports the possibility of exchanging data for statistical purposes with the statistical authorities of other countries. In the European Union, Regulation 223/2009 provides a legal framework for the exchange of confidential data between national statistical offices and central banks for statistical purposes. Some countries, however, may consider their national legislation more restrictive than the European Union legislation. This European Union legislation has provided a very helpful principle for improving the quality of monetary, financial and other economic statistics. It has also led to the closer integration of work, streamlining of data collection, reduction of costs and burden and a more effective exchange of knowledge. For this purpose, it would be very useful to have global and unique identifiers for identifying the same units across borders.

In some countries, statistical legislation prevents the exchange of individual data among producers of official statistics. In some cases, such exchange of individual data is allowed only without identifiers. This makes the exchange and linking of data sets challenging. In some countries, data exchange is allowed between a few organizations that have been explicitly mentioned in the legislation. National statistical legislation should allow the exchange of individual data among the producers of official statistics at the necessary level of detail, as recommended by the guidance.

At the national level, data also flow from other data providers towards the statistical system. The guidance recommends placing a legal obligation on administrative data providers to provide the necessary unit-level data to statistical offices for statistical purposes. This is considered a one-way data flow. Unit-level data are not provided to administrative data providers from the statistical authority, unless the organization produces official statistics in an independent organizational entity that belongs to the national statistical office. According to the guidance, statistical data shall not be used for any administrative decisions regarding the units. Other similarly discouraged uses of statistical data are listed in the guidance.

Currently, international data exchange for statistical purposes is not usually allowed or mentioned in the legal frameworks of countries. Outside the European Union (where the European Statistical System law defines the necessary concepts), it may be difficult to determine who is a producer of official statistics in the counterpart country, or who belongs to its national statistical office. Eurostat maintains and makes available lists of producers of European statistics.

Instead of trying to define all producers of official statistics of each country, the principle used in the common elements could apply to international data exchange. The producer of official statistics that authorizes access to or exchanges its confidential data, shall ensure that the recipient has the necessary legal framework in place for the full protection of confidential data. In addition, the guidance recommends that the exchange of data is voluntary, carried out exclusively for statistical purposes in the area of competence of each producer, must be authorized by the chief statisticians of the involved producers of official statistics and the conditions must be documented in mutually signed agreements. Lastly, the guidance notes that the agreements do not diminish the responsibility of the producer of official statistics to ensure the confidentiality of the data they exchange. Accordingly, it always requires careful consideration by the parties involved, but there should be no obstacles to data exchange if the legal frameworks fully guarantee statistical confidentiality.

Furthermore, typically data become subject to statistical confidentiality when entering the statistical system. Some countries have specific regulations listing unit-level data that are to be considered non-confidential. These may include business register variables, such as address, telephone, name, type of activity, number of staff and so forth of business units.

> Box 6.10 (*continued*)
>
> Even more often information on central and local public authorities, public health organizations, other public institutions and State-owned enterprises are exempt from statistical confidentiality. The Guidance on Modernizing Statistical Legislation also suggests including the possibility to the statistical law for the chief statistician to waive the protection of statistical confidentiality of data that are available to the public in accordance with other legislation, while considering the related quality and reliability issues. This could also enlarge the possibilities for the exchange of publicly available data among the statistical authorities of different countries.

6.6   Statistical sources

6.6.1   Economic census

6.107.   As discussed in chapter 2 and annex E1, an economic census can be conducted entirely independently of the SBR, or it can be SBR-assisted. It can involve the use of some administrative data, or it can involve only administrative data. The description in this section refers to traditional economic census where trained field enumerators seek out each physically recognizable place of business and collect the necessary information by direct interview and observation.

6.108.   Economic censuses are used in many developing countries and some developed countries, including the United States of America. An economic census is an intensive exercise and requires large inputs as well as resources. At the same time, it is undoubtedly a very useful instrument when a country is initiating an economic statistics programme. It provides benchmark data. In addition, in the past, especially in developing countries, it provided a well-established method for the initial construction of an SBR. In this latter respect, however, it has a number of drawbacks, as outlined below:

> ➤   In particular, it is a very resource-intensive exercise and requires large inputs of labour and time. Censuses, therefore, tend to be carried out infrequently, for example, once every five years.
>
> ➤   Intercensal updating of the SBR is thus required, which is itself costly. Once the census population of enterprises and establishments has been identified, additional data sources have to be used to maintain it. As some units have been found by enumeration and are not registered, administrative sources are not sufficient. There has to be continued enumeration in selected areas. This is particularly resource-intensive in developing countries where high levels of births and deaths of companies may occur.

6.109.   In addition to its high cost, the enumeration approach has the disadvantage of not being able to identify and document non-recognizable places of business, or enterprises without a fixed location, for example, web-based businesses, or individual entrepreneurs such as electricians and plumbers, providing services at locations other than their homes. In this sense it is important to search for administrative data which would cover such units or sources of other kinds, such as household surveys.

6.110.   In summary, economic censuses are not recommended as a means of establishing an SBR. Rather the converse is true; an economic census should draw its basic frame from the SBR, possibly supplementing this by an area sample. If, however, there is no reliable administrative source whatever, a periodic economic census is appropriate.

6.6.2   Feedback from enterprise and establishment surveys

6.111.   Feedback from enterprise surveys represents a vital mechanism for updating the SBR, as it provides information on changes in contact address, changes in the economic stratification characteristics, deaths and other factors. Feedback from enterprise surveys has the advantage that it is available at statistical unit level, that is, for establishments or enterprises. Accordingly, there should be close contact between survey staff and SBR staff to ensure that the SBR is updated with survey results. Survey staff should play an active role in SBR data quality management.

6.112.   Survey feedback has its limitations. An SBR updated exclusively by feedback from a traditional sample survey would have serious deficiencies. First, it would lack new units as surveys are not designed to find births. Second, the population of enterprises would not be fully maintained as feedback would be coming only from the sampled units.

6.113.   Furthermore, even for the sampled units, the use of survey feedback from sample surveys introduces the possibility of feedback bias, as SBR updates are provided only for the selected enterprises. There are no technical problems with using data for enterprises that have been sampled with certainty – typically the large ones. For medium-size and small enterprises, however, that are sampled with a probability less than one in repeating surveys, there is a potential for causing bias in future survey samples. In this situation, updates have to be very carefully applied.

6.114.   For example, suppose that when a particular quarterly survey is first conducted, the sample is found to contain 30 per cent dead enterprises (this is not an improbable figure). Furthermore, suppose that, based on this sample information, the dead enterprises are removed from the SBR and that the survey sample for the next quarter comprises the 70 per cent live units from the previous sample plus a replacement of the 30 per cent drawn afresh from the SBR. This new sample will contain about 9 per cent (30 per cent of 30 per cent) dead units. Thus, it will no longer be representative of the population of dead enterprises on the SBR, which is still nearly 30 per cent, assuming that the survey sample is a relatively small proportion of the population. There are proportionally too many live enterprises in the sample. If the survey weighting procedures do not take this into account (by making allowance for the dead enterprises that were originally found in the sample), the result will be an upward bias in the estimates. Furthermore, the bias will increasingly worsen with each survey repetition.[44]

6.115.   Several indicators for determining enterprise death should be taken into consideration while maintaining the SBR, samples and populations. These include the following:

> ‣ Information from other administrative sources, such as revoked licences, deregistration of VAT and other tax liabilities that are likely to be registered before business closures.

> ‣ Business demographic events at the establishment level often happen before enterprise closures are registered in the administrative registers and some of these events indicate business closures when lifted to the enterprise level.

> ‣ A statistical threshold for business activity is useful while defining business death, owing to the tendency of enterprises to have only minor (statistically insignificant) economic activity over longer periods of time before final administrative closures. The threshold should be defined on employment and turnover, but other economic variables could also be considered. Market concentration should be considered when the level of the threshold is set to ensure that the enterprises below the threshold only contribute to an insignificant part of the activity in economy and

[44] Example drawn from African Development Bank, *Guidelines for Building Statistical Business Registers in Africa* (Tunis, 2014), section 9.4.

individual industrial groups. In addition, the threshold procedure should take into account that the SBR may depend on different sources for different types of enterprises. For instance, some activities may not be liable for VAT and thus alternative sources are needed for turnover information.

6.6.3   SBR improvement surveys

6.116.   SBR updating information that cannot be obtained from surveys, or from the administrative sources on which the SBR is based, can be obtained by specific SBR improvement surveys (also termed "nature-of-business surveys", "proving surveys" or "SBR control surveys") conducted by SBR staff.

6.117.   Given limitations in resources, it is unlikely that all units in the SBR can be maintained equally. It is usually necessary to focus improvement surveys on specific strata to measure and improve coverage and quality.

6.118.   One possible strategy for keeping the SBR up to date is to conduct SBR sample surveys every year in which the biggest, the medium-sized and the smallest units are sampled to the levels of 100, 50 and 10 per cent, respectively. This practice helps keep values of the characteristics of the units up to date in an efficient way. Such surveys may also be specially designed to measure SBR accuracy, for example, to measure errors in classification by economic activity, or by size, or to estimate the proportion of falsely active units, as further discussed in chapter 10.

6.119.   In cases where there are risks that the sources (administrative or otherwise) used to build the SBR may miss businesses – in other words there is a risk of undercoverage – aerial samples may be drawn from which field listing activities may be undertaken and businesses missing from the SBR may be observed.

6.6.4   Profiling

6.120.   "Profiling" is the practice of using company accounts, often accompanied by interviews with senior enterprise officials, to build and define the structure of enterprises, mainly those involved in large complex enterprise groups. The resulting profiles are used to produce a reporting structure appropriate for the surveys conducted by the national statistical office. Profiling usually involves establishing contact with the enterprise being profiled to develop a good understanding of its structure. It is possible, however, to complete smaller profiles simply using published accounts.

6.121.   While profiling is not a primary source of data, it does provide valuable information on the larger and more complex enterprises that individually make a significant contribution to the country's GDP. It is especially important in identifying enterprise groups, as discussed in chapter 4.3.

6.122.   Profiling is often organized so that each individual SBR profiler has an assignment of large, complex enterprise groups for the reviewing and updating of which that profiler is responsible. When profilers react to signals from various sources on enterprise that may need updating this is referred to as "reactive profiling". Profilers also periodically perform "proactive profiling" to augment reactive profiling. For example, they may routinely arrange to reprofile every large enterprise once every three years. Since large enterprises can grow and change dynamically, considerable resources may be needed to maintain the accuracy and relevance of the SBR in this respect.

6.123.   The first step in profiling is to determine the criteria by which to identify the enterprises to be profiled. Profiling is usually focused on large complex enterprise groups with multiple activities for which survey reporting is difficult. After determining the

profiling criteria, a programme of regular updates over, say, a three-to-four year period, should also be established. There is a balance to be struck between the resources available and the amount of profiling that can be conducted, and this needs to be considered when determining the criteria. Resources should also be planned to deal with emerging issues that occur outside the routine reviews, namely, reactive profiling. For example, a major merger might require a profile to be conducted ahead of the scheduled regular review.

6.124.   The next step is to gather preparatory material. This should include records on all legal units within the enterprise group, the reporting history for surveys, and SBR data such as employment, turnover and classification, and so forth, for the enterprise group as currently defined. These data should be examined for consistency and to help identify reporting issues. As this process is labour-intensive, it is best to create a standard template for the data required and to automate the extraction of this information from the SBR.

6.125.   Further background information should be gathered by searching enterprise websites and examining annual accounts and reports. In simpler cases it may be possible to conduct a profile simply on the basis of this material, but for complex cases it is invariably necessary to meet with representatives from the enterprise.

6.126.   Following this preparation, contact is made with the controlling enterprise in the enterprise group (the global group head or global decision centre) and a visit arranged if necessary. Through discussion, it will be possible to identify the main trading (in other words, active) enterprises within the enterprise group and to agree on a mutually acceptable reporting structure covering these enterprises. The aim is to settle on a structure with enough detail for the statistical surveys, while minimizing the respondent burden for the enterprise group. Specially trained staff are required to negotiate with the enterprise as they will typically be talking to top management in the enterprise – the chief accountant or the company secretary or both. Ideally, these staff members will have a good understanding of the business processes and accounting practices, and also of the collection of economic data.

6.127.   Box 6.11 introduces a collaborative approach to profiling in Europe. More information on the methodology and practice of profiling may be found in annex F.

Box 6.11
European profiling: a collaborative approach

The restructuring of a number of important multinational groups in Europe over the past decade posed significant challenges to European statisticians in measuring global production. It was recognized that the "one enterprise equals one legal unit" principle no longer reflected the economic reality. In addition, the above events highlighted the importance of timely, high-quality, cross-border information on multinationals and their restructuring activities. Improved communication, coordination and data-sharing between the European national statistical offices were seen as a constructive way forward.

To better understand and monitor multinational groups in Europe, Eurostat and European member States have set up the EGR, a database on the largest European groups. This database includes the global decision centre of groups (which is the criterion for defining their country of residence) and all their legal units in the different countries.

In addition, as part of an overarching programme for the modernization of **European enterprise and trade statistics,** and as part of two consecutive European Statistical System projects, the European profiling methodology was officially agreed upon by the directors for business statistics of the European Statistical System countries in December 2017.

[45] In fact, because the perimeter of the large enterprise group changes every year, the profiling process should optimally be performed every year. These follow-ups offer the only means of keeping up to date and being able to calculate the most accurate statistics about enterprises.

[46] By December 2017, the following countries had tested the European profiling methodology over at least one year: Austria, Belgium, Bulgaria, Croatia, Czechia, Denmark, Estonia, Finland, France, Germany, Hungary, Iceland, Italy, Kingdom of the Netherlands, Latvia, Lithuania, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom.

Box 6.11 (*continued*)

The *European Business Profiling Recommendations Manual*, 2020 edition, was recently published by the dedicated Task Force on European profiling: https://ec.europa.eu/eurostat/documents/3859598/10479728/KS-GQ-20-002-EN-N.pdf.

European profiling is tested using individual standard grant agreements. An updated methodology was endorsed by the business statistics directors general at their December 2017 meeting. The updated methodology is characterized by the connection between the European and national profiling, designed to build synergies from the two processes for national users, the review of the profiling unit model, to correctly reflect the national enterprise picture and to link it to the global group structure of the group, and the review of the roles and responsibilities so that the profiling burden countries is more balanced among countries and the responsibility for decision-making on the resident statistical units rests with the respective national statistical office.

More than 300 multinational groups will have been manually profiled at least once[45] based on the European profiling methodology by 2019.[46]

Countries that have tested the collaborative process of European profiling have agreed on the following benefits:

➤ Significant benefits derive from the collaborative approach proposed.

➤ The top-down approach improves the correct understanding of the economic structure of multinational groups.

➤ Profiling leads to important improvements for the national business registers, especially as regards the additional legal units that can be discovered during the profiling process and the updating for the profiling reference year.

➤ Profiling fosters important long-term relations with respondents.

Annex F1 gives a brief introduction on the methodology of the European profiling.

6.7   Combining administrative and statistical sources

6.128.   To build a comprehensive SBR, a combination of administrative and statistical sources is recommended. Administrative sources identify enterprises, but may not include all of the required characteristics. Statistical sources do not identify new units but provide additional or more accurate characteristics. A strategy of using administrative and statistical sources in combination should be developed and employed. For example, administrative sources can be used to identify legal units and transform their data to form enterprises, while local units or establishments can be identified by a survey of the enterprises, as can characteristics missing from administrative sources.

6.129.   Using administrative and statistical data in combination is illustrated by the following examples:

➤   Suppose that an enterprise is thought to have only one establishment but employee data from an administrative source indicate that half of the workforce lives in an area far away from the identified establishment. This may be a signal that there is a second establishment, or that the employee data have been linked to the wrong enterprise.

➤   Suppose turnover from a VAT source does not correspond to the turnover for the same enterprise from an import and export administrative source. This may be a signal that some part of the enterprise is missing, or the links to one or other of the administrative sources are wrong.

> ▸ Suppose there is feedback from a survey indicating that the responding (legal) unit no longer has any economic activity. This may be a signal that the activity has taken over by another legal unit.

6.130.   Statistical sources can also be used to estimate missing characteristics. For example, if an administrative source only contains employee numbers and industry code, it may be possible to estimate turnover using turnover/employee value for similar units.

6.8   Record linkage in creation and maintenance of the SBR

6.131.   Usually, the main sources for the maintenance of the SBR are administrative data, such as data from taxation, social security or other administrative sources depending on their availability and adequacy. If these administrative sources have separate identification systems and if no identifier known to the SBR is provided, then record linkage methods have to be applied in order to be able to use the administrative data in the SBR.

6.132.   This section provides a guide to the theory and practice of record linkage. It is also important to note that there are commercial software products for record linkage. While these may be expensive, their use may be cost-effective compared to the investment required to build sufficient functionality in the national statistical office.

6.133.   Before any record linkage method is applied, it is necessary to analyse the administrative source to determine the quality of the source and whether it is useful for SBR purposes.

6.8.1   Basic approach

6.134.   "Record linkage" is the linking of a data record in one data source with one or more data records in another data source. In the case of the maintenance of the SBR it is the linking of one data record in the SBR with one or more records in the administrative data or registers, or vice versa. The challenge is to link records belonging to the same unit (whether this is a legal unit, enterprise, local unit, etc.) in different sources. If two records for the same unit are brought together this is called a match.[47]

6.135.   Where a common identifier is not available, identification of a match has to be based on a similarity measure that is computed using characteristics that are available both in the SBR and in the administrative source. The characteristics used are those that help identify a unit uniquely, typically name and address, and possibly other characteristics like legal form. The choice of characteristics may be different for different sources. Only a limited number of characteristics should be selected, namely, those that are most helpful in identifying units and least likely to be in error.

6.136.   It has to be taken into account that there may be errors in the data for units in either or both the administrative source and the SBR. For example, names may be misspelled or addresses may be out of date. Prior to matching, the quality of the characteristics of potential use for matching should be analysed in both sources.

6.137.   Another problem stems from the fact that the data in the administrative source may not be recorded in a format that is standardized and comparable with that used in the SBR. The length of the fields in the data records may be different, abbreviations may be used or one characteristic may be combined with another: for example, the legal form of a unit may be part of the name or in a separate field. Accordingly, parsing the format in the administrative source has to be done first and is a crucial step before matching methods can be applied. If not, the quality of the matching results will be lower than might otherwise be the case.

[47] An alternative is known as a "statistical match" and is defined as the linkage of records for similar units rather than for the same units. Statistical matching is ordinarily employed when the files being matched are probability samples with few or no units in common; thus, linkage for the same unit is not possible for most units. For more information, see Thomas N. Herzog, Fritz J. Scheuren and William E. Winkler, *Data Quality and Record Linkage Techniques* (New York, Springer, 2007), p. 81.

6.138.　It should also be mentioned that data linkage is quite computationally intensive: each data record in one data source needs to be compared with all the data records in the other source. Even where there are only a few hundreds of data records in each source, this will result in tens of thousands of comparisons. A matching run may therefore need hours of processing time. This can be drastically reduced if blocking methods are applied (as further described below).

6.8.2　Types of matching

6.139.　There are two basic types of matching: deterministic and probabilistic. The methods are very similar and are based on the computation of numerical values expressing the similarity of a record pair.

> ‣　In the case of deterministic matching, a similarity measure (or distance function) is defined. A unit of an SBR and a record of an administrative source are said to be a match if the two records have a high similarity for each element within the collection of identifying characteristics called the "match key".

> ‣　In probabilistic record linkage (formalized by Fellegi and Sunter)[48] a so-called "likelihood ratio" is computed. The value obtained is compared with two thresholds: a lower threshold, below which all values are defined as non-match, and an upper threshold, above which all values are defined as match. The grey area between lower and upper thresholds defines potential matches.

6.140.　Although probabilistic record linkage is mathematically elegant and has been successfully applied, its use in practice is limited by the fact that there is no implementation with a broad field of application in the standard statistical software packages. Deterministic record linkage is easier to implement and can result in equal or even better quality than probabilistic linkage.

6.8.3　Standardizing

6.141.　Matching should, as far as possible, be an automated process because of the large number of comparisons to be made. To achieve satisfying results with an automated matching algorithm the values of the characteristics being compared should not be formatted differently in the two sources. Before text characteristics are compared they must be standardized and parsed. Without standardization, many true matches would be erroneously designated as non-matches because the common identifying attributes would not have sufficient similarity due to differences in formatting. Standardization involves consistency of spelling, consistency of coding and elimination of obscure values.

6.142.　Standardization of text is essentially a statistical process. It is usually achieved by computing the frequency of all words in sources. If the frequency of a string (like "corp", "inc", "ltd") is very different in both registers, the string can be deleted from both sources, or abbreviated identically in both sources, or replaced by a synonym in at least in one source. An example of parsing is to ensure that all parts of a name are in the same order (e.g. family name, first name, title).

6.143.　Before starting automated matching, the contents of each text field must be analysed and the following questions must be answered:

> ‣　Are different languages used in the different sources? Are there combinations of letters that have a special phonetic value?

> ‣　Does the name field contain personal names, company names or both?

[48] Ivan Fellegi and Alan Sunter, "A theory for record linkage," *Journal of the American Statistical Association*, vol. 64, No. 328, 1969, pp. 1183–1210.

‣ Does the address field contain street names, city names or both?

‣ Are there any other peculiarities in a text field, such as the string "comp" preceding every company name?

6.144.   It is pertinent to consider how the data in the sources were obtained. Phonetic variations are likely to be found in data collected by phone, whereas typographical mistakes are probable in manually completed questionnaires. For standardizing purposes the following operations may be useful:

‣ Deleting blanks at the beginning or the end of a text field

‣ Converting lower-case characters to upper-case

‣ Deleting village name from the field for street name

‣ Converting special characters

‣ Converting Latin numerals to arabic numerals

‣ Moving the most crucial word to the beginning of the text field

‣ If corresponding text fields (such as company name) have different lengths, truncating the longer to the shorter length

6.8.4   Matching of text strings

6.145.   Given that the aim is to match two records (that do in fact refer to the same unit) by comparing text fields like name and address, if one or both of the two text fields has a typographical mistake, then the matching process will fail unless account is taken of the possibility of typographical errors. Thus, records and their textual characteristics (such as name and address) are considered as strings of alphanumeric characters and string comparator metrics are used to compare the two strings and to determine how much alike they are to each other. Usually the values of the metrics lie in the interval from zero to one, with one indicating perfect agreement and zero indicating high dissimilarity.

6.146.   There are many different possible string comparator metrics. Jaro[49] introduced a metric that accounts for the lengths of the two strings and for the types of errors that human beings typically make when writing or typing alphanumeric strings, like insertions, omissions and transpositions. ("Transposition" here means that a character from one string is in a different position on the other string. For example, in comparing "company" to "copmany", the "m" and "p" are transposed.) Winkler[50] augmented Jaro's metric by giving more influence to the characters at the beginning of a string. An extension of the Jaro metric is the Edit or Levenshtein metric. This is computed as the minimal number of edit operations (substitutions, imputations and deletions) which are necessary to convert one string to another string divided by the maximum of the two string lengths.

6.147.   Another important and widely used family of measures is based on N-grams. In this case it is not the characters themselves that are compared, but the sets of all N-grams of the strings. Usually either sets of all substrings of two (bigrams) or three (trigrams) successive characters are compared. The similarity measure is the total number of bigrams or trigrams that the strings have in common, divided by the average number of bigrams or trigrams in the two strings.

‣ For example: The bigrams of the string "Welcome" are "_W", "We", "el", "lc", "co", "om", "me" and "e_", for "Welkome" the fourth and fifth bigrams are different, being "lk" and "ko". Thus the measure is 6/8=0.75. As an option, the first and last bigrams containing the blank can also be left out, giving a measure of 4/6=0.67.

[49] Matthew Jaro,"UNIMATCH: a computer system for generalized record linkage under conditions of uncertainty. In *Proceedings of the May 16-18, 1972, Spring Joint Computer Conference.* New York: Association for Computing Machinery, pp. 523–530, 1972.

[50] William Winkler, "String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage", *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1990, pp. 354–359.

6.8.5　Blocking

6.148.　If in matching the SBR with an administrative source the numbers of record pairs is so high that it is not feasible to compute string comparator metrics for all possible pairs, then consideration should be given to blocking. "Blocking" is a method that reduces the number of pairs of records that are examined. In blocking, the files from the two sources are partitioned into mutually exclusive and exhaustive blocks. Comparisons are restricted to record pairs within the same blocks. Blocking is generally implemented by partitioning the two files based on the values of one or more characteristics. For example, if both files have a postal code field, the pairs to be compared might be restricted to those records whose postal codes agree. This would be an example of regional blocking. The advantage of blocking is that the number of comparisons may be greatly reduced. The disadvantage is that record pairs disagreeing on postal code are automatically classified as non-matches. Thus, if a record in one of the files has an erroneous postal code then it will never be matched even if there is a record for the same unit in the other file.

6.149.　This situation can be mitigated by iterating the matching process with a different blocking scheme, or even a sequence of blocking schemes. Suppose the first iteration used postal code as the sole blocking item and the second iteration used company name. Then failure to match due to an erroneous postal code would be somewhat offset by the possibility of matching during the second iteration. Further iterations could be made until the analyst felt that it was unlikely that matches would be missed because of errors in the blocking fields. In addition to subsequential blocking and multiple blocking there are other techniques that may be used, such as the sorted neighbourhood method, filtering, clustering and bigram indexing, all of which aim to reduce the search space of candidate record pairs.

6.8.6　Computing overall similarity measure

6.150.　The calculation of string comparator metrics for all compared pairs of records is usually done by a computer program. The metrics are computed separately and are combined afterwards, typically using a weighted sum, where the weights are defined according to the quality of the fields. The analyst has to define a lower and upper threshold for this overall similarity measure. Pairs of records with a value above the upper threshold are considered as matches, whereas pairs of records with a similarity measure below the lower threshold are considered as non-matches. When the metric lies between the lower and upper limit, that pair has to be clerically reviewed to determine if it is a match or non-match.

> ‣ The higher the upper threshold the smaller the number of false matches.
> ‣ The lower the lower threshold the smaller the number of missed matches.
> ‣ On the other hand, the further apart the thresholds the more manual work that is required.

6.151.　As manual work is time-consuming and therefore expensive, the choice of thresholds depends on balancing the impacts of false matches and of missed matches with the SBR resources available.

Box 6.12
Example of matching in the State of Palestine

PCBS is experienced in matching techniques using Hamming distance, Levenshtein distance and Jaccard distance matching algorithms and hybrid algorithms.

Despite the lack or complete absence of establishment IDs and enterprise IDs, the matching is carried out on the basis of three variables that are considered to be the most important for use in the automatic matching process: ID, commercial name and owner name.

A set of rules are applied to choose the values of the variables automatically. For example, the ID, commercial name and owner name of companies are taken from the data of the Ministry of Economic Affairs. The owner name of individual establishments is taken from VAT (as first priority) or from the municipality (second priority), and the address data are taken from the census, or the municipality (second priority), or the VAT file (third priority), and the phone numbers are taken from all sources without duplication.

Box 6.13
Importance of matching for the use of administrative data in order to achieve a system of linked registers: example of Denmark

It is always difficult to connect different registries. Some administrative sources do not hold ID numbers consistently over time, or do not have ID numbers at all. In these cases, one must consider whether to improve these input data by first adding a national statistical office ID, related to the source, in order for the unit to be maintained over time according to statistical needs and definitions. This will make an entry into the SBR easier and ensure a better quality. For each source, it is necessary to consider how best to apply it for the use of the SBR.

In Denmark all public registers use the same ID numbers for persons, legal and local units, dwellings and addresses. This practice makes it possible to link information across almost all registers by using the address as a unique ID across the three basic registers: Central Person Register (CPR), Building and Dwelling Register (BDR), and Central Business Register (CBR). Their relationship is shown in figure 6.3.

Figure 6.3
Three basic registers in Denmark
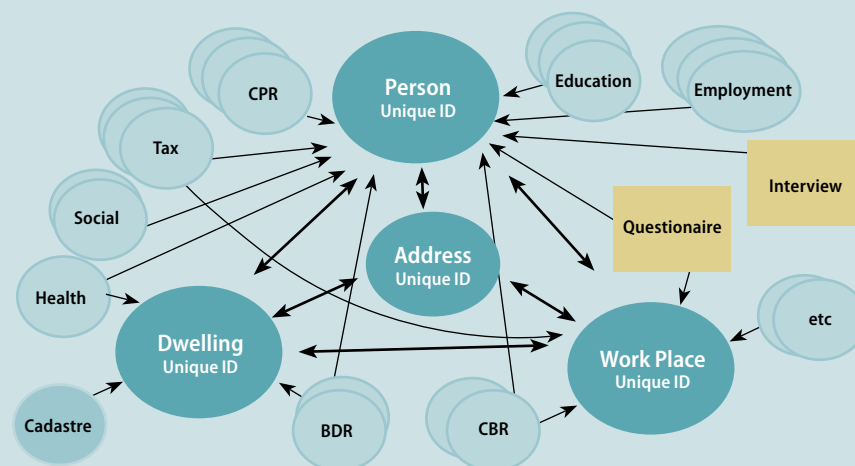
Box 6.13 (*continued*)

One of the most important data items for quality assurance and linking registers is the address. It is therefore important that national statistical offices work to create the best possible address register – either for themselves or as a basic register for wider use in society. For example, Denmark has for many years had an official address register. By 2017 it became mandatory for all public authorities to use the Common Address Register as the basis for their addresses. In addition to enhancing the quality of the individual registers, it also ensures easier coordination between registers. In Denmark, there is free access to the Common Address Register, which means that commercial providers also use data from this source.

Figure 6.4
Addresses without common keys
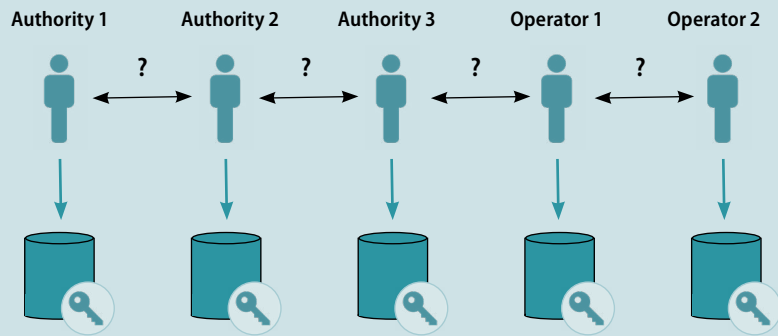
Figure 6.5
Addresses with common keys
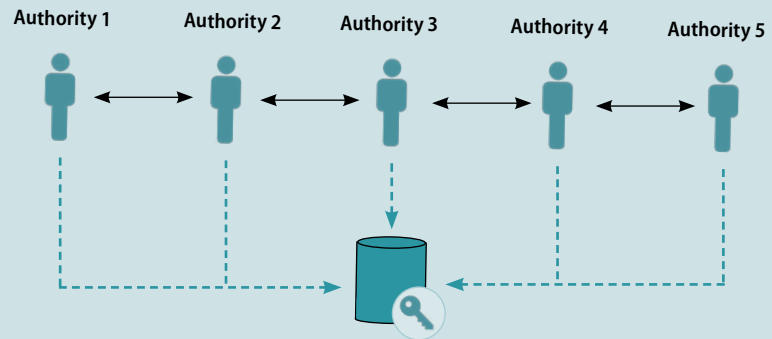
Figure 6.6
Addresses as a common reference
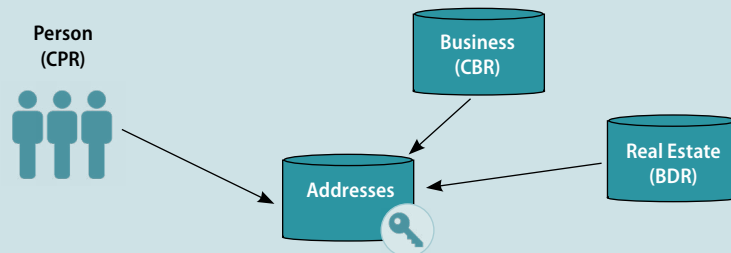
Box 6.13 (*continued*)

With the address as a common key linking the three basic registers: CPR, BDR, and CBR, the national statistical office of Denmark can ensure that they always refer to the same point on the ground, thereby facilitating the combination of data from different registers and the geographical breakdown of statistical results.

This also means that businesses do not have to register their address in each and every satellite register as long as there is a unique identification for them in any one of the three basic registers. This results in fewer data updates and fewer errors in the administrative sources.

6.9   Other data sources

6.152.   In addition to the data sources discussed in previous sections, there are commercial sources and, added to these, new sources are emerging, associated with increased computer power and online form filling, or the filing of data in administrative systems, as outlined in the following paragraphs.

Telephone directories

6.153.   Telephone directories or special listings prepared by telephone companies can be useful in adding or confirming SBR data. They should not be used as sources of new enterprises.

Payroll, taxation and accounting service providers

6.154.   Payroll, taxation and accounting service providers provide enterprises with services that involve some or all of such processes as paying an enterprise's staff, making returns to the taxation authorities on its behalf and managing its accounts. It may be possible for the national statistical office to build agreements with such service providers and their client enterprises that allow the service providers to provide data for the enterprises directly to the national statistical office. This saves the enterprises from the burden of survey questionnaires and provides faster and more efficient data flow to the national statistical office. Identification of the relevant enterprises is required. It may still be necessary, however, to contact the enterprises directly to obtain information that the service providers cannot provide, for example economic activity classification.

Internet search

6.155.   The Internet may be useful as a supplementary source of data, but, at the current time, cannot be reliably used for identifying new enterprises. It can, however, provide information on the economic activity, on the production profile, on up-to-date addresses, and on other elements, so is becoming an important source.

Commercial data providers

6.156.   There are a number of commercial enterprises that provide global, regional and domestic company information. These organizations serve businesses, enabling them to make informed decisions in such areas as credit, marketing and others. Their

data are also valuable for maintaining an SBR, in particular by providing information on enterprise group structures.

6.157.   Data from commercial data providers are, by definition, based on publicly accessible information and could also be obtained directly by the national statistical office. At the same time, the equivalent collection, matching and processing within the national statistical office would require significant resources and may well prove to be more expensive than purchasing these tailor-made data.

Big data

6.158.   Big data are defined as large, possibly unstructured data sets that are potentially available in real time. The opportunities for using big data are still being developed. So far no practical experience in their use in the SBR has been identified. In the future, however, big data may reveal opportunities for developing additional sources for the SBR. Accessing big data sets, which are often owned by private businesses such as mobile telephone operators or Internet providers, might pose a challenge. Structuring and editing the data might also require substantial investment but be worthwhile if accompanied by significant benefits in terms of coverage and timeliness

Box 6.14 provides some examples from national statistical offices on the exploration of new data sources and new technologies.

Box 6.14
Continuous pursuit of new or additional data sources: China, Colombia, Mongolia and Tunisia

China

Owing to significant differences between the sharing economy and the traditional economic model, the existing statistical methods and data-collection channels are not well adapted to the census needs of the sharing economy.

The National Bureau of Statistics in China has published a report on a pilot system for the statistical monitoring of new industrial and business forms and models, recognizing the sharing economy as an important part of the new economy. The report recommends that new questions be added to regular statistical surveys. It is more difficult to set up special surveys for different trading entities for sharing economic transactions. Statistical indicators related to the sharing economy are added to the enterprise statistical statements, in combination with the current enterprise surveys, so as to obtain statistical indicators reflecting the participation of various subjects in sharing economic processes.

The National Burau of Statistics is endeavouring to make full use of big data technology to explore new data sources. It will further improve the statistics of network trading platforms and explore the use of big data technology to expand data sources.

Colombia

The records from private companies such as electricity suppliers in Colombia are a source of interest for updating and maintaining the SBR. Currently, information is available in the Energy Billing Master Framework, which contains records associated with non-residential units. In the exploratory stage of these data, however, it became clear that there is no unique identifier which can link the records with those already existing in the Directory of Companies, so that checks are still being carried out to determine whether or not they can be included. It should also be noted that private companies provide the information in a non-periodic manner, with the result that updates from these types of data sources are not regular.

Box 6.14 (*continued*)

The National Administrative Department of Statistics is continuously searching for new data sources for the SBR, which will make it possible to enrich the stored information, providing additional features that complement the framework on which the users base their work.

The incorporation of new sources constitutes a relevant process in the regular updating of the SBR, making necessary the inclusion of social data such as those reported by the Ministry of Health through the integrated contribution payment form, which is linked to the social security system. Not all sources, however, are integrated in the same way to the operation of the SBR, because in some cases the information does not have a recognized identification number. It is thus necessary to use geocoding and georeferencing as an integrating element.

Working with the SBR makes it possible to measure demographic behaviour in each of the sectors of the economy, based on the use of the registers that have been stored. It is not only important to expand the coverage of the SBR but also to guarantee the conservation of historical data, in order to identify, analyse and even predict behaviours of interest.

The Department is also working on a big data project that uses web-scraping techniques to supplement the primary data source of the Statistical Business Register. In this project, extracted data from web pages are used to improve coverage and to update data on the identification and location variables of enterprises.

The project focuses on automating the processes of searching and collecting data on the Internet that would serve as a source for complementing and updating the basic information of the companies registered in the SBR, as well as a source for the search and identification of new companies.

Based on the design of the project, the data collection is carried out by web robots (computer programs that collect information from the web automatically) using automatic navigation on a few selected websites that can provide the most relevant commercial information in the economic industries of tourism and lodging businesses in Colombia.

Actual implementation of the project consists of four phases:

- Phase 1: Analysis, which includes such activities as the selection of economic sector, the definition of variables to complement, the update and selection of data source, and the analysis of successful cases.
- Phase 2: Design, which involves such activities as approval process, selection of programming languages, and definition of conceptual model.
- Phase 3: Development, which involves programming and integration scripts.
- Phase 4: Deployment, which includes activities to verify results and to make adjustments to programs and scripts

Key activities performed under the project include:

- A compilation of international references on the use of web-scraping techniques carried out by National Statistical Offices was analysed, as a complementary form for the collection of information for official and national interest research frameworks.
- Two web robots were developed for the extraction of information published on two selected websites dedicated to tourism and lodge businesses.
- These two web robots captured and structured information on the description of the establishments, and also on their contact and georeferenced location in Colombia from the pages of the two selected websites.
- Data downloaded from the two websites included more than 8,000 and 24,000 businesses respectively.

Box 6.14 (*continued*)

> ▸ The respective methodological document on the processes was generated during project development.

Based on the results of this project, it was concluded that web-scraping techniques can provide additional data sources for SBRs and improve their quality. While the project focused mainly on tourism and lodge businesses, more research needs to be done for other industries in the economy.

At the time of writing, two more robots are being developed to explore new sectors of the economy and to improve the coverage of the SBR.

Mongolia

The national statistical office of Mongolia is planning to further integrate its SBR with the database of the electronic payment system (e-billing system) that contains data on VAT and on quarterly income tax of entities and individuals, along with data on the general government sector institutions that are registered in the civil service agencies and the Ministry of Finance. Furthermore, the national statistical office of Mongolia plans to develop an application that can automatically update the SBR and cross-check with these databases.

The application will make the estimation processes much faster and more efficient, enabling data to be disseminated at a faster pace and become more comprehensible for users, enterprises and government organizations. Consequently, data quality, survey coverage and the accuracy of the estimated indicators based on statistical units will all be improved.

Tunisia

In order to improve the quality of the SBR in Tunisia, INS opted to diversify the sources of the SBR update, which includes administrative sources (DGI, CNSS, Customs), and business surveys (national economic survey, employment and wages survey, micro enterprise survey, short-term surveys and others).

The updating of the SBR is done mainly on the basis of the administrative files. Given the problems in these sources, however (lack of detail on the activity, late or no update for addresses and activities, among others), INS uses the information received from the field when visiting companies in the framework of economic surveys. This information is used to update the main variables of the SBR, in particular for the address, the activity, the turnover, the number of employees and others.

Chapter 7
Maintenance of the statistical business register

7.1   Introduction

7.1.   Whereas chapter 6 describes the data sources used in SBR construction, this chapter discusses maintenance of the various types of statistical units and their characteristics. The key objective of maintenance is to update the coverage and content of the SBR, taking into account continuity and stability rules, according to a well-defined calendar, and in as timely a fashion as the information sources allow. The basic aim is to provide economic surveys with sampling frames that are accurate and as up-to-date as possible.

7.2.   Section 7.2 presents key aspects of the SBR maintenance strategy. It discusses the data sources, in particular administrative sources, and how to deal with conflicting information. It describes the practice of allocating enterprises to maintenance groups for maintenance purposes, with different procedures for each group according to the size and complexity of the enterprises that it includes. It discusses the timing of maintenance in relation to the production of survey frames.

7.3.   The various types of changes that can occur to statistical units in the live register, and how they are handled, are described in sections 7.3 and 7.4. The former section deals with births, deaths and changes in the linkages of statistical units. The discussion is presented in terms of a typology of demographic events. The latter section discusses all other changes, namely, changes in the characteristics of statistical units.

7.4.   Continuity rules for determining whether a statistical unit is deemed to have continued despite significant changes (e.g. of ownership, size, economic activity or location), or to have died and been replaced by another unit, are discussed in section 7.5, as are stability rules (sometimes called resistance rules) that restrict the speed with which changes of characteristics are applied in order to inhibit unwanted oscillations in values.

7.5.   The final section introduces the distinction between changes that actually occur in the real economic world and corrections of errors caused by inadequacies in SBR procedures, and it discusses the sources and treatment of errors.

7.6.   The primary references used in developing this chapter are the 2010 Eurostat Business Registers Recommendations Manual (chaps. 12–18), the 2012 African Development Bank Guidelines for Building Statistical Business Registers in Africa (chap. 10), and the 2007 *Eurostat – OECD Manual on Business Demography Statistics* (chap. 4).

7.2   Statistical business register maintenance strategy

7.7.   As observed by the African Development Bank, "a country's economy is constantly changing – new businesses are formed, existing businesses merge, change production activities or location, go bankrupt, etc. To ensure that enterprises (and other

[51] African Development Bank, *Guidelines for Building Statistical Business Registers in Africa,* chap. 5.7.

standard statistical units) in the SBR remain aligned with and representative of legal units and their productive activities, these changes have to be detected and the SBR has to be correspondingly updated. This process is referred to as SBR maintenance".[51]

7.8.   Maintenance involves constant updating of the SBR in terms of coverage and content. New statistical units must be identified and recorded without overcoverage, undercoverage or duplication. Changes in the values of the characteristics of existing statistical units must be identified and recorded. The dates of changes should be known and recorded. Ideally, the information should be collected just once.

7.9.   The maintenance strategy should take into account the following aspects:

> The sources of information – administrative sources, economic censuses, feedback from surveys and SBR improvement surveys – and rules for dealing with conflicting information

> The maintenance groups into which statistical units in the SBR are divided for the efficiency and cost-effectiveness of maintenance procedures

> The time dimension – timing of the updates, the need for continuity and stability rules, and use of the historical register

7.2.1   Use of administrative sources

7.10.   In terms of number of updates to statistical units, administrative sources are the major contributors to SBR maintenance. It is recommended that one administrative source be identified as the primary source for construction and maintenance purposes. The source chosen should be one that provides the best possible balance of coverage, content, timeliness and accuracy. In using additional (secondary) administrative sources, care must be taken that no duplication of units arises. Duplication can most easily be avoided if the secondary sources use the same unit identification scheme as the primary source, or have no overlap with the primary source. An example of the latter case is where the primary source is a tax register and the secondary source is a list of government departments. In the absence of common identification scheme, or coverage that is known to be complementary, records in a secondary source must be matched to those in the primary source and links established between records referring to the same unit, thus avoiding creating duplicates of that unit.

7.2.2   Use of feedback from surveys

7.11.   Feedback of changes in frame data from previously conducted economic surveys is a valuable source of updates to statistical units. By design, the first few questions asked by any economic survey should constitute a check of the values of basic characteristics (such as name, address, contact information, and activity status) of the unit recorded in the survey frame. Inactive or dead units should be identified. Subsequent questions may collect updated versions of economic activity codes and size measures.

7.12.   There are no conceptual problems in updating the SBR with data for any statistical unit that has been sampled with certainty. For medium-size and small enterprises that have been sampled with probability less than one, there is potential, however, for causing bias in future survey samples, as described in section 6.6.2. Thus, it is recommended that feedback from surveys about characteristics (such as economic activity and size) that are used for sampling statistical units should not be used to update units in the SBR that were not selected with certainty. Instead, the information can be used trigger further maintenance operations and/or can be taken into account in future repetitions of the survey, for example by not trying to contact units that are known to be dead.

7.2.3 Use of SBR improvement surveys

7.13.   The objectives of SBR improvement surveys (also called "nature-of-business surveys", "SBR proving surveys" and "SBR control surveys") are to verify the current values of key characteristics and to obtain missing values. Such surveys are conducted by the SBR on a continually repeated basis, with sample sizes matching the SBR resources available. The aim in selecting the samples is to have maximum beneficial impact in terms of SBR quality improvement. There is no focus on estimation. The total sample size over a year is determined by taking into account the frequency with which changes occur in the economic world, the quality of data received from the administrative sources, the quality of data currently in the SBR, and the resources available.

7.2.4 Use of census data

7.14.   Where there is insufficient availability of administrative data, countries might be forced to use the results of an economic census for the updating of the SBR. Economic censuses, however, are undertaken only every five or ten years. The process of data collection, editing and tabulation usually takes a considerable time. Census results might be available only some years after the reference period. Thus, the data collected are no longer representative for the current period. This time lag is specifically addressed in the two country examples below.

Box 7.1
Time lag of census data: Egypt and Indonesia

Egypt

As the country's economic census is conducted once every five years, CAPMAS cannot rely on it alone to update the SBR, given the large number of changes during those five years that will not be captured by the economic census. CAPMAS can, however, benefit from the economic census in the area of business demography, since the demography of the business population is represented by data on:

- ‣ The active population of enterprises
- ‣ Their birth
- ‣ Their survival (followed up to five years after birth)
- ‣ Their death

CAPMAS calculates the survival rates of the establishments by comparing data with those of the successive economic census.

Indonesia

The economic census in Indonesia is conducted once every 10 years, the last one being in May 2016. The economic census aims to determine the profile of businesses in Indonesia, which can then be used for planning macroeconomic activities, and the resulting data can be used as a reference for the subsequent surveys. Given the dynamic nature of economic data, within 10 years there may be many changes, including in business profile and characteristics, economic level and structure, competitiveness, the structure of expenditure and income, capital, prospects and business constraints. The current time lag between economic censuses is therefore relatively long, rendering it impossible to reflect routine economic fluctuations.

7.2.5   Profiling

7.15.   As discussed in section 6.6.4, the aim of profiling is to produce an appropriate statistical and reporting structure for large and complex enterprises. Profiling can be reactive, namely in response to a signal from any of the various sources that there has been a change. It can also be proactive, typically involving a reprofiling programme in which every enterprise that is either large, complex, or likely to experience material change is reprofiled every so many years.

7.2.6   Dealing with conflicting information

7.16.   As evidenced in the previous paragraphs, the SBR is updated from a range of sources, several of which may provide values for the same unit and characteristic. When this happens, there may be a conflict in the values. This raises the question of which source and value to use. The answer requires a thorough understanding of each source – the methods of data collection and validation, the time of collection, and the relative importance assigned to that characteristic by the source. This understanding then allows the sources to be prioritized, either in respect of all units, or differently for various categories of units.

7.17.   Once priorities have been determined for each characteristic, the next step is to apply them to SBR updating procedures. This may be done in several ways. The most reliable method is to record the source and date of recording of the value of a characteristic and to ensure that the updating algorithms specify which combinations of source and date can result in the value being overwritten subsequently. Date is important because it allows for the possibility that a new value from a lower priority source can overwrite the existing value from a higher priority source if the information is significantly more recent.

7.2.7   Maintenance groups

7.18.   The amount of maintenance effort devoted to a statistical unit should be commensurate with its size and potential impact upon published statistics, and should take into account its propensity to change and the sources of updating information. Thus, statistical units should be divided by size and potential impact and propensity to change into maintenance groups, each of which is subject to a particular set of updating procedures. Three examples are provided in box 7.2.

Box 7.2

Maintenance groups in Australia, South Africa and Switzerland

Australia

Enterprises maintained by the Australian Bureau of Statistics (ABS): These enterprises are maintained by the SBR staff through profiling, the SBR survey, business survey feedback and ad hoc investigations, as appropriate in accordance with their size, complexity and propensity to change. A trigger-based approach to profiling has been introduced to better reflect real-world changes as they occur.

Enterprises maintained by the Australian Tax Office: These enterprises are currently maintained entirely by taxation data from the Australian Tax Office. The capability to include ABS-sourced information has, however, recently been introduced into the SBR with a view to exercising this option to improve data quality in the future.

Box 7.2 (*continued*)

South Africa

Enterprises subject to profiling: This set of enterprises is maintained by profiling operations. Indications of changes from administrative sources or from business surveys are used as signals indicating the need for reprofiling. The enterprises are not included in SBR improvement surveys, nor are they subject to ad hoc investigations.

All other enterprises: All other enterprises are maintained using updating information from administrative sources and economic survey feedback, which is supplemented as needed by information gathered by SBR improvement surveys and ad hoc investigations.

Switzerland

Enterprises subject to profiling: All enterprises with more than 10 local units, or more than 100 employees, are included in the profiling group. The profiling staff contact them directly every quarter to gather structural information and employment data.

Enterprises with only one local unit, including small businesses: These are maintained using updating information from administrative sources and economic survey feedback.

Enterprises subject to light profiling: All other enterprises are included in the light profiling group. The profiling staff contact them annually by Internet survey to gather structural information and employment data.

This maintenance strategy is dictated by the need to determine employment for all local units as the regional distribution of employment is an essential component of Swiss business statistics. All enterprises with only one local unit are included in the SBR improvement survey, which also enables quality control of activity codes for the entire SBR.

7.2.8 Updating schedule

7.19.   The SBR is updated from several sources of information, each having a given periodicity, coverage and content. This must be taken into account in building a coherent maintenance system. The schedule for maintenance of the SBR and production of frozen frames should be coordinated with the supplies of data from administrative and statistical sources, with the production of frames for the surveys using the SBR, and with the publication (if any) of data directly from the SBR. The schedule indicates to users the likely variations in SBR coverage and content over time.

7.20.   The SBR should be updated on a daily basis. Potential updates cannot be allowed to accumulate. For this reason, the SBR is often referred to as the "live register". The timing of the supplies of updating data should be discussed and coordinated with the sources. The production of frozen frames can then be undertaken in a systematic and controlled manner, taking into account the whole economic statistics programme and, in particular, the conflicting demands from structural surveys (that want the most up-to-date information available for the reference period for which the survey is being conducted) and subannual surveys (which want stability across reference periods so that the effects of maintenance do not appear as artefacts in survey estimates of change).

7.21.   A frozen frame is never directly updated. A preliminary version of a new frozen frame is compared with the previous version. Unexpected changes are analysed. If they turn out to be due to errors in the SBR, then these errors are corrected and the production of the frozen frame is repeated.

<div style="background-color:#d5e4e4; padding:1em;">

Box 7.3
Statistical business register reports in Canada and Switzerland

Canada

Statistics Canada produces SBR reports and also allows survey staff to see preliminary versions of their survey frames after key update processes have been completed each month.

Statistics Canada's Business Register Journal

In order to keep track of the reasons for the changes, the Business Register Journal is used to store the reason for the change when a profiler, SBR user or updater performs an update. The Journal entry is created automatically with a link to the unit or units being updated when the update is performed. The entry contains auto-generated scripts of the event or update and allows the profiler or updater to complement this information with additional information. All the information in the Journal is additional to the update log in which the following information is kept: previous information or value, new information or value, date and source of update, and for some events, the effective date, namely, the date on which the event actually occurred in the real world.

Switzerland

In the Federal Statistical Office (FSO) of Switzerland, a working group on economic classifications produces reports for the users on important updates in economic activity classification.

</div>

7.22.   Staff of surveys using the SBR as the source of their survey frames should be aware of the SBR production calendar and should be kept informed about any particularly significant updates in SBR coverage or content occurring since production of the previous frozen frame that may significantly affect their survey frames.

7.2.9   Historical register

7.23.   As described in chapter 2, the historical register is an SBR output that enables the reconstruction of the history of the units. It supports:

> ‣ Obtaining information about births, deaths and continuity of units.

> ‣ Obtaining information about size development (how fast a unit or group of units is growing, through what kind of economic activity).

> ‣ Analysing changes in the characteristics of units, for example, in location, judicial and financial links, and economic activity.

> ‣ Reconstructing the state of the SBR as of a past date or reference period.

7.24.   As a prerequisite, the dates of all demographic events and changes in values of characteristics should be recorded. If this is not possible, dates for core changes should be recorded.

7.25.   In principle, a historical register can be derived and maintained in three different ways:

> ‣ The SBR is considered to be established afresh at the beginning of each year. During the year, all changes that each unit undergoes and the dates and reasons for those changes are recorded. At the end of the year, the SBR is copied and stored. Together with the change records this copy becomes a historical register.

‣ The snapshots of the SBR taken at regular intervals (typically quarterly or annually) provide a series of pictures of the structure of the register. From each of these snapshots, the numbers of units and their characteristics at a specific point in time can be obtained.

‣ The SBR records and maintains complete information about changes as they occur on a continuous basis. Each change is date-stamped with the date on which it was recorded in the SBR and the date on which it occurred in the real world. This allows populations of units and their characteristics to be constructed for any point in time. It also provides information about delays in registering changes. Storing the reasons for the changes is also useful but is much more difficult. As a compromise, some changes can be automatically identified and categorized on the basis of a set of automated rules and procedures.

Box 7.4
Statistical business register maintenance strategies: Colombia and Mauritius

Colombia

Overview of the SBR programme

The SBR provides strategic information for the generation of frameworks based on statistical operations; for this reason, it contains the administrative records provided by the official information sources of the business population and business demographics. The SBR coverage includes all companies (natural and legal persons) that carry out an economic activity and contribute to GDP.

The Statistical Directory has significantly increased the volume of information from approximately 4,460,000 records in 2013 to 13,800,000 records in 2017. The large volumes of stored data provide a solid source of information for economic analysis, strengthening national accounts, meeting research demands and updating sampling frames.

Administrative registers and surveys are the main sources for updating the SBR, providing information on identification variables, location, stratification and demographic component. The quality of the information is manifested by the accuracy of the data, the frequency of the updates and the coherence of the processes, aspects in which there are standardized processes and defined metrics.

Updating process

Considering the dynamics of the legal units, continuous maintenance of the SBR is required. For this purpose, the updating process is defined as a constant cycle that seeks to incorporate and update information from the identified data sources. The phases of information management, preparation, incorporation and dissemination are carried out according to the frequency of availability of the data sources (administrative records and surveys), which, in the case of Colombia, corresponds to annual and semi-annual periods. In each of the aforementioned stages, information quality processes are provided, which are defined and documented in operation manuals.

Figure 7.1
Statistical business register update phases
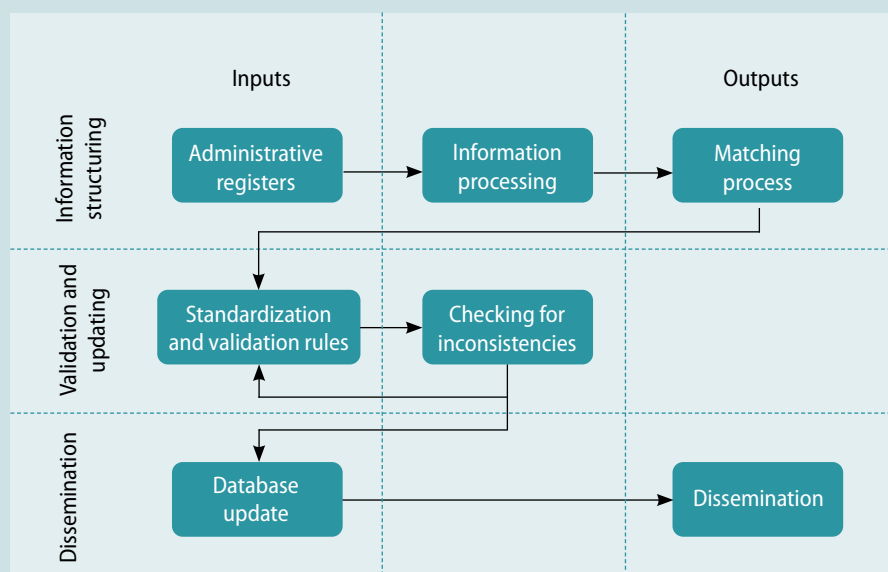
Maintenance strategies

The process of updating the SBR involves the structuring phase of the information aimed at identifying and correcting errors in the data provided by the sources through the administrative records, which is important for improving the quality and consistency of the stored variables. When multitemporal analyses of economic phenomena are carried out, however, the results may be inconsistent. The use of character similarity algorithms to make the comparisons between registers provides an alternative means of finding duplicate records between the bases delivered by the providers, with a view to structuring only the information that has undergone any change between periods. This in addition can speed up the preparation process, ensuring the quality of the new information prior to its incorporation. Validation and verification scripts executed periodically provide a diagnosis of the structure and completeness of the data; likewise, the analyses generated through the data cubes offer support for the process of checking the coherence of the information.

Unique identifier

The legal units that are included in the SBR have a unique identification number that is assigned by the Tax Department for tax control. Its use is generalized by all public and private entities and this allows information integration. However, in addition to the identification number, the SBR implements a code in order to carry out at system level the maintenance of the records, speeding up the search processes.

Administrative sources

The administrative registers and surveys supplied by the sources provide a set of data with information of interest for the consolidation and maintenance of the SBR. Currently there are registers supplied by the Superintendence of Corporations, the National Confederation of Chambers of Commerce and the Ministry of Health and Social Protection, and surveys of the National Administrative Department of Statistics, which are conducted periodically. The continuous search for new sources review in detail the consistency, relevance, reliability and coverage of the information before including the new data source into the updating process.

Figure 7.1 (*continued*)

Web-scraping techniques are used to evaluate potential suppliers that report information on the website associated with the economic sector of interest, taking into account criteria of reliability and data quality. The process begins with the programming of a robot that enters the defined websites to collect information related to identification data, location and size of legal units. Next, processes of cleaning and standardizing data are carried out in order to obtain greater precision in the crossing of sources; lastly, non-matching records are identified in each of the analysed websites to produce a unified result. Considering that the output of generated information constitutes a new provider, the data preparation phase is carried out before incorporating the records into the Statistical Directory of Companies.

Mauritius

Administrative data are acquired from the selected administrative sources and stored within the SBR database. Updates of or changes in the administrative data since the previous occasion on which they were acquired are processed, for the most part automatically. These changes may result in the creation of new statistical units within the SBR and in changes to existing units. In addition, they may generate queries for manual resolution.

It is to be noted that input data from all sources are treated in essentially the same way in the SBR. Records are copied with minimal change from the source into a staging area where they are checked for consistency against the record layout, range checks and other editing rules. Records that fail the checks remain in the staging area until they are individually and manually corrected or deleted. After manual correction they are again subjected to the consistency check.

Records that pass all checks are copied into an area where consistent data from previous cycles for that source are stored. These records are also used as the source of updates to the SBR statistical units. This is their primary function. The actual updates to the statistical units depend upon the processing specifications that have been supplied by the business registration unit in the form of item-by-item precedence rules. In essence, a precedence rule determines in what circumstances a new piece of data will cause the data item value of a statistical unit to be updated. Precedence rules are currently expressed simply as a priority ordering of the various possible sources.

For example, for an item such as turnover, data from the VAT source have precedence over data from the Corporate and Business Registration Department (CBRD). Employment data from Statistics Mauritius annual surveys (businesses employing 10 or more persons) will have precedence over employment data obtained from the Ministry of Social Security; status (birth/death) information from CBRD will have precedence over other sources, although timeliness is an issue because of lengthy legal, administrative procedures and non-reporting by small entities.

A simple priority ordering based on source will not be fully satisfactory in all circumstances. In particular, the timeliness of data plays a part in their credibility and this needs to be taken into account. It should be possible to specify that a very recent value from a lower priority source is more credible than a much older value from a higher priority source. Consideration is being given to the ways of doing this without rendering the precedence rules so complicated that it becomes overly hard for the users to determine why certain updates have occurred or not occurred.

7.3   Handling births, deaths and linkage changes of statistical units[52]

7.3.1   Demographic events

7.26.   A demographic event is defined as an event that has an impact on the existence of a statistical unit, or on links between statistical units.

7.27.   A demographic event is based on changes in the existence of production factors, or in their distribution, within and among statistical units. It may involve the continuity (survival) of a unit over time or its discontinuity (death). It may also be accompanied by changes to the values of certain characteristics, such as size or type of economic activity.

7.28.   Figure 7.2 presents a general typology of demographic events relating to enterprises and enterprise groups, based on two distinct types of events:

▸   Events involving existential changes, such as the emergence or disappearance of combinations of production factors.

▸   Events involving distributional changes, such as changes in the distribution of production factors between enterprises.

7.29.   The typology reflects the importance of the enterprise as the core statistical unit.

7.3.2   Demographic events, continuity and information sources for enterprises

Information sources

7.30.   Most events are detected after the reception of signals, typically changes in the administrative data that are used to maintain the SBR. Additional information may be needed, either directly collected or from other sources, to decide what updates to enterprises should be made based on these signals. The updating procedures may well differentiate between micro, small, medium and large enterprises.

7.31.   Creations and cessations of administrative units do not necessarily result in births and deaths, respectively, of enterprises. An enterprise may be born, or may die, without a change in the legal unit(s) that own(s) and/or control(s) it. Detection of enterprise deaths is more difficult than detection of enterprise births as administrative sources are quick to add new units but slow to indicate those that have ceased to have economic activity.

Figure 7.2
Typology of demographic events [53]

1		**Changes of existence of combinations of production factors**
	1.1	Emergence of combinations of production factors (birth)
	1.2	Disappearance of combinations of production factors (death)
2		**Changes in the distribution of production factors**
	2.1	Redistribution of the production factors within one enterprise
		2.1.1 Redistribution of production factors across local units
	2.2	Redistribution of the production factors of more than one enterprise

Figure 7.2 (continued)

	2.2.1	Merger, takeover (concentration of enterprises)
	2.2.2	Split, break (deconcentration of enterprises)
	2.2.3	Transfer of production factors between enterprises
	2.2.4	Enterprise restructuring
2.3		Redistribution of the production factors within one enterprise group
	2.3.1	Redistribution of production factors across local units of more than one enterprise
	2.3.2	Redistribution of production factors across enterprises
2.4		Redistribution of the production factors of more than one enterprise group
	2.4.1	Concentration of enterprise groups
	2.4.2	Deconcentration of enterprise groups
	2.4.3	Transfer of production factors between enterprise groups
	2.4.4	Enterprise group restructuring

Definition of continuity

7.32.   An enterprise is considered to continue through a demographic event if its production factors continue. It is considered to die (be discontinued) if its production factors are discontinued. In practice, continuity is interpreted as meaning that at least two of the three following characteristics remain essentially the same: controlling legal unit, economic activity and location.

7.33.   An enterprise is deemed to continue if it resumes its seasonal activities, or resumes its activities within a specified period after stopping them. Such a period needs on the one hand to be short enough so that the SBR is appropriately up-to-date, on the other hand appropriately long enough so that the empirical evidence that a unit stopped its activity not just temporarily is verified with due consideration also of the time lag in the process of recording in the administrative registers. In Europe, 24 months is assumed to be an adequate period.

Birth

7.34.   The birth of an enterprise is the creation of a combination of production factors with the restriction that no other enterprise is involved in the event. Handling a birth in the SBR means creating a new enterprise unit, in other words, an enterprise unit with a new identification number. A birth is typically detected by a signal in the form of the appearance of a new unit in an administrative source. As noted above, however, not every new administrative unit results in the birth of an enterprise.

Death

7.35.   The death of an enterprise is the dissolution of a combination of production factors. Handling a death in the SBR means giving the enterprise the activity status of "dead", also described "ceasing" the unit. Although this does not actually involve erasing the unit from the database, this is also sometimes referred to as deleting the unit. Deaths are typically detected through survey activities and through the disappearance of units from an administrative source.

Change of ownership

7.36.   A change of ownership occurs where a new legal unit takes over the production activities of an existing enterprise. This event by itself does not affect the continuity of the enterprise and, therefore, should not cause the birth or death of an enterprise. Typically, however, a change in ownership results in the death of one administrative unit and birth of another. Thus, in practice, for small enterprises that are maintained entirely by data from an administrative source, a change of ownership may well result in the death of an enterprise and the birth of another one (resulting from the birth and death of the corresponding administrative units). For a large enterprise, a change of ownership may be detected through profiling and thus the continuity of the enterprise can be detected and the enterprise maintained.

Restructuring within an enterprise

7.37.   Restructuring within an enterprise, for example, the creation of a local unit, does not affect the continuity of the enterprise, although it may be associated with changes in characteristics such as size or economic activity.

Concentration (merger, takeover)

7.38.   A concentration is an event where there is more than one legal unit before the event and only one legal unit after the event. More precisely:

- ➤   If all the legal units before the event lose their identity, the event is a merger.
- ➤   If one of them retains its identity, the event is a takeover.

7.39.   In principle, mergers and takeovers do not necessarily involve the deaths or births of enterprises as such, though some characteristics (such as size and economic activity) may change. Quite often, however, an SBR is designed in such a way that enterprises are tightly linked to legal units. Thus, in the absence of additional information from profiling:

- ➤   In the case of a merger, all the enterprises corresponding to the legal units that existed before the event are ceased and a new enterprise is created.
- ➤   In the case of a takeover, the enterprise corresponding to the legal unit that takes over the other legal units continues and the enterprises corresponding to the legal units that were taken over are ceased.

Deconcentration (split-off, break-up)

7.40.   A deconcentration is an event involving one legal unit before and more than one legal unit after the event. More precisely:

- ➤   In a break-up, the original legal unit dies; all the legal units after the event are new.
- ➤   In a split-off, the original legal unit continues; the other legal units are new.

7.41.   In principle, break-ups and split-offs do not necessarily involve the deaths or births of enterprises as such, though some characteristics (such as size and economic activity) may change. As noted above, however, quite often the SBR is designed in such a way that enterprises are tightly linked to legal units. Thus, in the absence of profiling information, the births and deaths of enterprises follow those of the legal units.

Reconstruction of demographic events from statistical SBR information

7.42.   The cause of a new enterprise record can be a birth, a merger, a break-up or a split-off. The cause of an enterprise being given the activity status of dead can be a death, a merger, a takeover or a break-up. To reconstruct a demographic event more precisely, it is necessary to determine all the enterprises involved and to have recorded their links to one another over time and the dates of changes in those links.

7.3.3   Demographic events, continuity and information sources for local units

Definition of continuity

7.43.   The continuity of a local unit is defined in terms of continuity of its location and production factors, with an emphasis on those production factors that can be readily identified at the level of the local unit, namely, land, buildings, and employment.

- ➤ If the location remains the same, the criteria for local unit continuity are, first, the continuity of production factors, in particular of employment; and, second, the continuity of the enterprise to which the local unit belongs. A local unit that does not change location is deemed not to continue, in other words, to lose its identity, if at least two of the following three factors change: the enterprise identity, the principal activity, or at least 50 per cent of the employment.
- ➤ If a local unit changes location within a region (defined at the local level), it is deemed to continue, in other words, to retain its identity, if none of the three factors mentioned above change. Otherwise, it is deemed to be discontinued.
- ➤ If a local unit moves outside the region it is deemed to be discontinued.

7.44.   A local unit is deemed to continue if it resumes its seasonal activities or resumes its activities within a specified period after stopping them. In Europe 24 months is assumed to be an adequate period.

7.45.   The birth of an enterprise that comprises only one local unit implies the birth of a local unit.

7.46.   The death of such an enterprise implies the death of a local unit. A local unit belonging to an enterprise with only one local unit may, however, be discontinued and be followed by the creation of different local unit while the enterprise continues.

Birth or death

7.47.   The birth of a local unit is the creation of a (partial) combination of production factors at a geographically identified place. A death is their dissolution.

7.48.   Births and deaths are handled by creating new local unit records and marking existing ones as dead, respectively.

Transfer

7.49.   A local unit may be deemed to continue and be transferred from one enterprise to another. This may be handled by deletion of the link between the local unit and the enterprise to which it belonged before the event and the creation of a link between the local unit and the enterprise to which it belongs after the event.

7.3.4   Demographic events, continuity, information sources – enterprise groups

Typology of demographic events for enterprise groups

7.50.   In the context of enterprise groups, the focus is on events reflecting concentrations (mergers, takeovers) and restructuring, and their impacts, notably on employment at both national and international levels. Births and deaths are far less frequent for enterprise groups than for enterprises.

7.51.   There are certain differences in the recording of events and in the allocation of identity numbers according to type of enterprise group (multinational group, truncated group and all-resident group). This is in order to ensure coherent handling of the continuity of multinational enterprise groups (MNEs) and their truncated parts.

7.52.   The categories of the general typology of demographic events for enterprise groups differ from those for enterprises in two respects:

> ‣   In the event of the death of an enterprise group, it loses its identity, whereas redistribution of production factors does not necessarily involve identity loss. This is because the (real) death of an enterprise group is the cessation of all control links, direct or indirect, between the legal units of which the enterprise group consists. The legal units become independent again or cease to exist. No other enterprise group is involved. The death applies only to all-resident and multinational groups, not to truncated groups, which may cease to exist through other events.

> ‣   The numbers of enterprise groups both before and after the event are decisive. For instance, the birth of an enterprise group and the concentration of two existing enterprise groups differ, among other things, in the number of enterprise groups involved. This is because births and deaths of enterprise groups are not in one-to-one correspondence with the creation and deletion, respectively, of identity numbers. There are two reasons for this. First, births and deaths are events (concerning the real, observable world) whereas the creation and deletion of identity numbers are SBR updates. Second, a birth or a death involves only one enterprise group, whereas other events, such as mergers, that involve many enterprise groups may also result in the creation and deletion of identity numbers.

7.53.   The Eurostat *Business Registers Recommendations Manual* (2010 edition) restricts births and deaths to the involvement of only one enterprise group. This convention is chosen because it is in line with the terminology of users who are interested in questions such as "What is the effect of enterprise group births on a certain economic activity?" If the term "birth" is used in the sense of these types of questions, enterprise groups emerging, for instance, from mergers or split-offs are not included, although their impact on concentration (or deconcentration) is relevant. Further details may be found in box 7.5.

Existential changes (births and deaths)

7.54.   The birth of an enterprise group is the establishing of a control link or links, direct or indirect, between two or more independent legal units, where no control link previously existed, and no other enterprise group is involved. Temporary links

of less than one year are not taken into account. Birth applies only to all-resident and multinational groups, not to truncated groups, which are created through other events.

7.55.   The death of an enterprise group is the cessation of all control links, direct or indirect, between the legal units of which the enterprise group consists. The legal units become independent again or cease to exist. No other enterprise group is involved. Death applies only to all-resident and multinational groups, not to truncated groups, which cease to exist through other events.

Changes within an enterprise group

7.56.   Changes may be divided into three categories:

> ‣   Change of global group head (controlling unit)
>
> ‣   Creation or deletion of a truncated group
>
> ‣   Restructuring within an enterprise group

7.57.   A change of global group head involves the replacement of the controlling legal unit by another legal unit. Recording of the event should include the date when it takes place. This is very important for the statistics of foreign affiliates compiled according to the country of the ultimate controlling unit. Although the legal unit changes, this event should not in itself affect the continuity of the enterprise group. It does not result in the birth or death of an enterprise group. Depending, however, on the sources used to create the enterprise group, the enterprise group may be defined according to the controlling legal unit and the source may issue a new identifier in the event of a change of controlling unit.

7.58.   Creations and deletions of truncated enterprise groups are important special cases of restructuring within an MNE. In general, they do not affect the continuity of the multinational group, but they change its structure and possibly its characteristics. An important case is when there are several control links crossing national borders leading to seemingly more than one truncated group within the same multinational group. There are two possible options for handling truncated groups and their heads. The different groups of legal units thus formed within a national territory may be combined into a single truncated group or each group may be defined as a separate truncated group. As previously discussed in section 4.3, the first option is recommended.

Concentration (merger, takeover)

7.59.   As in the case of enterprises, if both enterprise groups lose their identity, the event is called a merger. If one of them retains its identity, it is called a takeover. It should also be emphasized that a takeover may lead to changes in some characteristics of the enterprise group that retains its identity. For instance, it may enter a different size class or have a different principal economic activity.

7.60.   Mergers and takeovers only apply to all-resident and multinational groups. There are no similar types of events for truncated groups. If two seemingly separate parts of a truncated group that belong to the same multinational group become merged under a national group head, this is not a real merger. From a global viewpoint it is restructuring within the enterprise group. While it is desirable to track and record such events, this can be difficult at national level and has to be coordinated at global level.

Deconcentration (split-off, break-up)

7.61.   A break-up is an event, where an enterprise group is divided in such a way that none of the resulting enterprises groups retains the identity of the original group. A split-off is an event where one of the resulting enterprise groups retains its identity.

7.62.   As with mergers and takeovers, break-ups and split-offs apply only to all-resident and multinational groups. There are no similar types of events for truncated groups.

Complex restructuring

7.63.   Complex restructuring involves more than one enterprise group before and afterwards. An example is the transfer of an enterprise, parts of enterprises, or a number of enterprises between two or more enterprise groups. If the continuity of employment within an enterprise group is used as a criterion for the continuity, this may lead to discontinuity of an enterprise group even if no change of global group head has been detected.

7.64.   Restructuring may also affect the characteristics of the enterprise groups that retain their identities. For example, their principal economic activity or size class may change.

7.65.   The phenomenon of complex restructuring has been prominent in the economic and financial press for many years, so there is certainly a high demand for such information.

Box 7.5
Typology of demographic events: example from the European Union

The Business Registers Recommendations Manual (2010 Edition) of the European Union provides a typology developed specifically for enterprises. It is based primarily on changes in the existence of production factors and their distribution within and among business organizations, while taking into account enterprise structure in terms of organizational units of production, economic activities, locations, and legal and financial links.

Events are separated into two groups:

 ▸ Those involving existential changes, namely, the emergence or disappearance of combinations of production factors
 ▸ Those involving distribution changes, namely, changes in the distribution of production factors between units

As indicated in the following table (where n implies 2 or more) existential changes (births and deaths) involve a transition from no enterprise to one enterprise, or vice versa, whereas changes in the distribution of production factors require that at least one enterprise is present both before and after the event.

Event	Real, observable world		Business register	
	Number of enterprises before the event	Number of enterprises after the event	Number of creations	Number of deletions
Birth	-	1	1	-
Death	1	-	-	1
Change of ownership	1	1	-	-
Merger	*n*	1	1	*n*

Box 7.5 (*continued*)				
Takeover	*n*	1	-	*n*-1
Break-up	1	*n*	*n*	1
Split-off	1	*n*	*n*-1	-
Creation of a joint venture	*n*	*n*+1	1	-
Cessation of a joint venture	*n*	*n*-1	-	1
Restructuring within an enterprise	1	1	-	-
Restructuring within an enterprise group	*n*	*n*	0 or more	0 or more
Change of group	1	1	-	-
Complex restructuring	*n*	*n*	0 or more	0 or more

Translation of events into SBR updates

Births and deaths

7.66.   In the SBR, an enterprise group birth is represented by creating a new identity number; and a death by marking the enterprise group as dead. A more difficult problem is posed by determining the date of birth. In the case of all-resident groups, the date of birth is, in principle, the date when the control link is established between two or more legal units. In practice, as the smallest all-resident enterprise groups are not monitored, the date of birth is more likely to be the date when the enterprise group is created in the SBR. The same considerations apply to birth in an MNE.

7.67.   For a truncated enterprise group, the date of birth should, in principle, be the date on which the first unit is established in the national territory. In practice, the date is more likely the date of registration of a corresponding administrative unit with an administrative source.

Changes within an enterprise group

7.68.   Changes may be divided into three categories:

> ‣   Change of global group head (controlling unit)
> ‣   Creation or deletion of a truncated group
> ‣   Restructuring within an enterprise group

7.69.   Changes in group head or restructuring within an enterprise group are events that should not have a direct impact on the demographic characteristics of the group. They are, however, important signals that may trigger further investigation. In the case of an MNE, they may have considerable impact at national level, even the creation or ceasing of a truncated group, and they may be reflected through changes in relationships or characteristics recorded in the SBR.

7.70.   As previously noted, if there are several parts of an MNE in the national territory, these should be combined and recorded as a single truncated group. This avoids the creation and deletion of truncated groups when restructuring takes place within the MNE. On the other hand, bringing together truncated groups that have several, seemingly independent parts renders SBR maintenance more complicated at the national level.

Mergers and takeovers

7.71.   In the case of a merger, all the identity numbers of the enterprise groups existing prior to the event are ceased and an identity number is created for the emerging enterprise group.

7.72.   In the case of a takeover, the enterprise group that takes over the other group retains its identity number, so no creation takes place. The enterprise groups that have been taken over are marked as ceased. The date on which the change is considered to have taken place is the date when both parties have accepted the merger or takeover or when it has been approved by the competition authorities (if this is required).

7.73.   The creation and ceasing of identity numbers apply in the same way for all-resident and multinational enterprise groups. Truncated enterprise groups simply follow the events at multinational group level.

Break-ups and split-offs

7.74.   In the event of a break-up, identity numbers are created for all the enterprise groups existing after the event and the identity number of the original group is marked as ceased. In the event of a split-off, new identity numbers are assigned to the split-off enterprise groups.

7.75.   The creation and ceasing of identity numbers apply in the same way for all-resident and MNEs. The truncated groups simply follow the events at multinational group level.

Restructuring involving more than one enterprise group

7.76.   Complex restructuring may entail any number of SBR creations and deletions of different statistical units and, in practice, is very difficult to manage. For example, in the European Union, restructuring of multinational groups should be coordinated by the EGR, whereas restructuring of all-resident groups takes place at the national level.

Information sources

7.77.   In practice, most events are detected as a result of signals from administrative or commercial sources (which may themselves depend upon public administrative sources) or from statistical sources. In some cases, additional data are needed, either collected directly or obtained from other sources.

7.78.   In all cases, cost-efficiency must be considered. Thus, the SBR updating policy may differentiate between small and large enterprise groups. Small enterprise groups may be updated based on administrative sources, while for large enterprise groups, profiling is usually necessary.

7.79.   Although administrative sources differ widely from country to country, it is generally the case that creations and cessations in administrative sources, or – and especially – in commercial, sources, do not necessarily result in SBR updates as they tend to follow entirely different continuity rules. Nevertheless, administrative information about new enterprise groups can be a good start for determining appropriate statistical updates.

Reconstruction of demographic events from SBR information

7.80.   The cause of an SBR creation of a new enterprise group can be birth, merger, break-up, split-off, or restructuring. The cause of an SBR ceasing of an enterprise group can be death, merger, takeover, break-up, or restructuring. Accordingly, mere registration of creations and cessations in the SBR does not enable derivation of precise data on the underlying demographic events. More information is needed.

7.81.   To reconstruct a demographic event, the enterprise groups involved have to be identified. In the cases of concentration, deconcentration and restructuring it is necessary to have links over time between the enterprise groups involved. For example, in the case of a merger, the original enterprise groups must be linked to the emerging group and, in the case of a takeover, the enterprise group that is taken over must be linked to the continuing group. If such links are recorded with dates, all events can be reconstructed. This implies the need for a historical register as discussed previously.

Continuity in relation to the definition and use of the enterprise group

Continuity in theory

7.82.   A general definition of enterprise group continuity based on its definition, uses and central strategic role concerning the units it comprises may be phrased like this:

> ▸   If the enterprise group has continuity in its decision-making on its overall policy for production, sales and profits, financial management and taxation, and has a centralized strategy concerning the units that it comprises, it is considered to be continuing.

7.83.   Production factors are important for enterprise groups, although less so than for enterprises. An enterprise group may often use subcontracting for much of its production, thus the physical production factors play a much smaller role, while the roles of strategic planning, research and development and intangibles are very important. This adds to the complexity as the physical production factors are generally much easier to take into account than intangibles.

Continuity in practice: basic rules

7.84.   The discussion on continuity rules here concerns both multinational groups and all-resident groups, although the rules may need to be applied differently, depending on the availability of information. The continuity of truncated groups in the SBR should follow the continuity of the MNE to which they belong, in addition to the events discussed above. When information on the continuity of the multinational group is available, the continuity of the truncated group can be decided accordingly. Only other events (creation, deletion and so forth concerning the truncated groups themselves need to be handled separately).

7.85.   Continuity of multinational groups should be coordinated at international level. For example, a procedure in the EGR allocates an identity number in a unique format to each multinational group (a national statistical office can of course also assign an additional national identity number). In the case of discontinuity, this identity number is changed. An indication of discontinuity should come from the country of the group's decision centre when this occurs in Europe (which is the case for the majority of groups operating in Europe), or when there is a European group head. When the decision centre is outside Europe, a decision regarding continuity has to be based on the available information from, first, the country (if any) where a European group head is located; second, the commercial sources; and, third, published information, for instance in the European Union Industrial R&D Investment Scoreboard. The identity numbers and their changes are communicated to the national statistical offices where the enterprise group operates.

7.86.   Continuity of all-resident groups may be decided according to cost-efficiency on the basis of different methods but according to agreed rules. For large groups, profiling may be used for the decision. For small groups, automated procedures based on administrative and SBR information can be applied.

7.87.   Only annual substantial changes are taken into account, not long-term developments. Changes are found by comparing the situation at a certain time in the year to the same time in the previous year. As in the case of large enterprises, sudden substantial changes in large groups are rare. One such event could be a possible change of group head. On the other hand, smaller changes within the group, or restructuring between groups, are very frequent. They may also occur subannually, in which case they may not necessarily be taken into account in the annual monitoring of the group.

7.88.   A key factor to take into account is how the enterprise group itself sees its continuity. There must be strong reasons to go against the group's own opinion, in particular because discontinuity among the largest groups is rare.

7.89.   In determining continuity the following factors are taken into account

    1. Global group head, composition and location(s)

    (a)   Global group head

    (b)   Enterprises which belong to the group

    (c)   Main location(s)

    2. Economic activity

    (a)   Principal activity according to ISIC

    (b)   Employment

    (c)   Intangible assets

Use of the continuity components

7.90.   If all these factors change, discontinuity is obvious. In other cases, the changes need to be identified and weighted in order to make a decision. An important case, in particular for smaller groups, is when a new group head appears and changes the whole profile of the group, both principal activity and employment, and possibly intangible assets. This event should be considered as the creation of a new group, even if the group head remains the same.

7.91.   The main factors, namely 1 (a), 1 (b), 2 (a) and 2 (b), may provide a suitable basis for the creation of an algorithm on which an automated decision about continuity can be based.

Change of global group head

7.92.   The convention is that if there are no changes other than a change of global group head, there is deemed to be continuity of the enterprise group. Thus, change in group head is not in itself sufficient for discontinuity. It is, however, an important indication of the need to check if there are changes in the other continuity factors.

Change of principal economic activity

7.93.   The principal activity may be assumed to be positively correlated with the continuity of the production factors. This criterion, however, is probably not relevant for multi-activity enterprise groups, where relatively small changes to one part of the group may cause a major change to the economic activity code of the group as a whole. In addition, a gradual shift in activities may (after respecting an appropriate stability rule) eventually cause a change in economic activity code, resulting in a reclassification of the enterprise group. In such cases, there is continuity and the change of activity may be disregarded.

7.94.   To use the principal economic activity in determining continuity requires following the mix of economic activities of the group members in terms of value added, if available, or, if not, turnover. Employment may also be used, in particular in cases where no reliable turnover is available for each enterprise, only consolidated turnover for a VAT group. It has also been found that employment is also generally more stable over time than turnover.

7.95.   For multi-activity groups, use of principal economic activity in determining continuity is very sensitive to the definition of the principal activity (especially at four-digit level) and also to whether some predefined threshold (such as 50 per cent) is used. Thus, principal activity as a single criterion is not a good approach. For all-resident groups in particular, however, a sudden change of principal economic activity when combined with some other changes, for example of group head or employment, is a strong indication of discontinuity. At global level, principal economic activity as a criterion is more difficult to apply.

Change of employment

7.96.   If the employment of the legal units in the group remains much the same from one year to the next, the group may be considered as continuing.

7.97.   Continuity of the national group head is strongly correlated with continuity of employment, but as it may change more easily, employment is a more suitable measure.

7.98.   Using employment is a very practical continuity measure as it can even be totally automated, but it has weaknesses. One weakness is that changes may occur in the legal units due to restructuring within the group. The main weakness is that it is not clear how the approach can work in a global framework due to the lack of availability of the necessary data and their timing.

Change of main location

7.99.   The locations of the global group head and head office have a strong impact on the enterprises in a group, on the continuity of locations where the research and development are carried out, and on where the actual production takes place. Unlike in the case of enterprises, however, practical rules on how to use main locations as a criterion for enterprise group level continuity are difficult to define and have not been used.

7.4   Handling changes in characteristics of statistical units

7.4.1   Changes in identification and contact characteristics

7.100.   Identification characteristics enable the identification of units and their linkage to other units in the SBR and in other sources. They enable the tracking of units over time, which is particularly important when there is conflicting information about the dates when changes occurred.

7.101.   As previously noted, changes in contact characteristics should normally be reflected in SBRs as soon as they are detected as these data are used for mailing survey forms and for geographical analyses of SBR data. Changes to a unit do not necessarily imply changes to linked units.

7.4.2   Changes in economic stratification characteristics

7.102.   Economic stratification characteristics are used in determining the probabilities of units being included in survey samples. It is therefore important to consider whether:

> ➤ To update these characteristics as soon as new information is received and risk increased volatility in survey populations, samples and the resulting estimates, or

> ➤ To hold the updates back until the point in the annual cycle of surveys when the impact of changes on the consistency of results is minimal, or

> ➤ To subject the updates to stability rules (also termed "resistance rules") that inhibit changes, as further discussed in section 7.5.

7.103.   Changes in legal form, for example from unincorporated enterprise to a limited corporation, may affect the consistency of survey populations and samples over time. Cross-checking name and legal form can help in identifying potential problems.

7.104.   Guidance on coding, and changes of coding, of the institutional sector of legal units and enterprises is provided in annex B2.

7.105.   When the SBR is used to coordinate surveys, for example through the use of permanent random numbers, the impact of updates on coordination should also be assessed.

7.106.   Changes to size characteristics (persons employed, number of employees, number of employees in FTE and turnover) have significant effects on survey samples.

7.107.   Large and complex units require more attention because changes to such units have more impact on statistics and there is an increased incentive to ensure the changes are correct. Potential updates should be checked by comparing data from a range of sources to see there is consistency, or by contacting the units concerned directly. As large units are often included in surveys with certainty (that is, with probability 1), survey feedback is the most common source of updates. In addition, users of the SBR – in particular, managers of economic surveys and the national accounts – should be consulted and informed about updates of large and complex units. Contributions to production are typically the basis for the procedures and algorithms used to derive principle economic activity when an enterprise is involved in more than one economic activity, and legal form when the enterprise involves more than one legal unit.

7.108.   Potential updates to small units have less impact on the resulting statistics and for that reason the procedures for their maintenance can be largely based on automatic updating from administrative sources, without the need to contact units to confirm changes.

7.109.   Changes to geographical location codes for enterprises and local units have to be consistent with the corresponding continuity rules for these units.

7.110.   Information received on changes in the various activities of a particular unit may prompt a change to the principal or secondary activity code for that unit. These changes may be sudden, for example, due to a change in management policy for the business concerned, or gradual, where the balance of activities has shifted over time. In all cases, use of stability rules is recommended in order to suppress movements that are no more than temporary phenomena or statistical artefacts. In particular:

> ➤ For large units, the stability rule may involve a case-to-case decision by a group of experts.

▸ For small units, automated rules may be used, for example, a change in principal activity may be made only after evidence for it has been sustained for two years.

7.111. For businesses engaged in certain activities, such as tourism, there is likely to be a seasonal pattern to the monthly or quarterly numbers of persons employed and turnover. Data for monthly or quarterly reference periods are useful for detecting short-term trends and turning points; they may, however, give a misleading view of the sizes of units from an annual perspective. It is therefore recommended that averages over an annual period are also held for such characteristics and used for sampling purposes.

7.112. When stratification characteristics are updated, it is useful to record at least the date of the change, the source of the new value, and the previous value. These are valuable in helping assess the quality of the values of the characteristics and in auditing change processes.

Box 7.6

Maintenance of the economic activity code in the Swiss Business and Enterprise Register

The economic activities of the units in the Swiss Business and Enterprise Register (SBER) are defined by the General Classification of Economic Activities (NOGA, www.bfs.admin. ch/bfs/en/home/statistics/industry-services/nomenclatures/noga.html). NOGA, which is the Swiss national version of NACE Rev. 2, counts 794 positions at the lowest – six-digit – level (as compared to the 615 positions of NACE Rev. 2 at its lowest – four-digit – level). The split in national positions mainly applies to wholesale and retail trade, monetary intermediation and human health activities.

Codification system

Each unit registered in SBER has a NOGA code. The codification of new units is based on the following two steps:

1. *Provisional codification*

If a new unit is registered in the Register of Commerce, it receives a provisional NOGA code given by a private FSO partner.

For new units announced by other sources (especially administrative registers), the provisional codification is directly done by SBER staff.

2. *Final codification*

The provisional code for enterprises consisting of one local unit becomes final within the scope of a quarterly based survey to update the Business and Enterprise Register (ERST), in which all new enterprises in SBER are asked, among other information, about their economic activity.

In the case of multi-establishment enterprises, units are directly integrated into the profiling activities.

Update of the activity code

NOGA codes are used for statistical purposes exclusively and may only be modified if the change is justified. This can happen in two situations:

▸ If the company changes its main economic activity
▸ If the company has been incorrectly coded

Box 7.6 (*continued*)

The following sources are used for the updating the activity code:

- ‣ Administrative and branch registers
- ‣ Statistical surveys
- ‣ Profiling and NOGA quality survey
- ‣ Private partnerships
- ‣ Hotline
- ‣ Monitoring

Administrative registers

Various administrative sources (such as the commercial register, the VAT register), and also branch registers (register of medical professions, lawyer and notary register), are regularly updated. Their information is constantly integrated in SBER, enabling the continuous updating of the economic activity codes.

Statistical surveys

Specific statistical surveys (census of the agricultural enterprise structure, hospital statistics, pension fund statistics, bank sector statistics and others) provide information about the economic activities of the enterprises. Some sample surveys also contain additional questions on the activity of the responding enterprises.

Profiling and the NOGA quality survey

The enterprise profiling and the NOGA quality survey ask companies directly about their economic activity.

Profiling

Profiling activities are based on a close partnership between companies and the FSO.

This cooperation between the FSO and large multi-establishment enterprises, which deliver quarterly information to the FSO, makes it possible to keep up-to-date data on the company structures in SBER and their economic activity codes and reduce the burden on the enterprises considerably.

NOGA quality survey

For enterprises consisting of one local unit, SBER has installed a control system regarding their economic activities which ensures a systematic update of the NOGA code. This system has a quarterly sample survey as its core.

In a first step, all enterprises which already indicate their activity in their address (such as restaurants, hairdressers and others), and also enterprises with a website are discarded from the sampling. Enterprises which are already integrated in other surveys (such as hotels, banks and others) are also removed as well from the sampling.

The remaining enterprises are then divided into four classes, based on the principle: "the bigger the enterprise, the more frequent the check of the activity code". Large enterprises, with more than 100 employees, are surveyed every two years, mid-size enterprises every three years, small enterprises every four years and enterprises with only one or two employees every five years (see fig. 7.3).

Figure 7.3
Cycles of the NOGA quality survey

Class N°1	Class N°2	Class N°3	Class N°4
>100 employees	10-99 employees	3-9 employees	1-2 employees
2 years	3 years	4 years	5 years
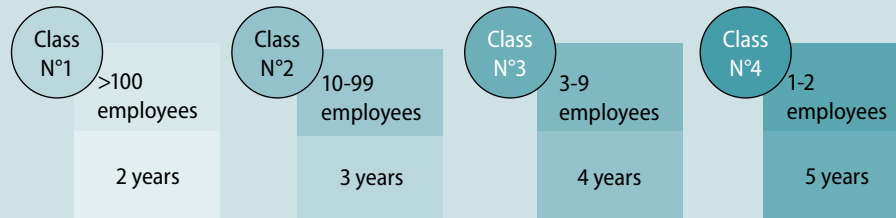
Private partnerships

A partnership has been established with external data sellers that regularly interview businesses, among other on their economic activity. This information is handed over to the FSO and can lead to a change of the NOGA code of the businesses in question.

Hotline and help desk

The FSO unit responsible for economic classifications offers a hotline for enterprises, taking care of questions about the classification in general, but also about specific economic activity codes. They also receive regular requests for updating the NOGA code of enterprises which will lead to an actualization or improvement of the NOGA code, if the request is justified.

Monitoring

All NOGA codes change at the six-digit level for companies with 50 or more employees and at the two-digit level for companies with 20–49 employees are weekly monitored.

The newly attributed codes are all analysed by the classification staff and confirmed or rejected.

The coding tool for classifications, known as "KUBB", also includes the explanatory notes of the different versions of NOGA, the explanatory notes of the statistical classification of products by activity (CPA) nomenclature. The above-mentioned tool facilitates the association of an economic activity to a NOGA code.

In order to guarantee a certain coding quality, the equal handling of difficult cases and international comparability, the FSO has developed a coding manual and keeps it updated. This document is mainly based on Eurostat recommendations, cases discussed in the web discussion forum of Eurostat (CIRCABC) and inputs observed in Switzerland. All these rules are discussed in advance and approved by the statistical partners of FSO. Training is continuous and based on mistakes encountered and difficult areas of coding to keep employees and partners up to date.

7.5　SBR maintenance procedures

Legal and operational units

7.113.　Updating information about legal and operational units linked to large enterprises is obtained through profiling.

Administrative units

7.114. Information about administrative units is obtained from the corresponding administrative sources. The data received should be stored without change in the SBR. This leads to the following outcomes:

> ‣ In the case of a new administrative unit appearing in incoming administrative data, a corresponding new administrative unit is recorded in the SBR.

> ‣ In the case of changes in values of the characteristics of administrative units appearing in the incoming administrative data, the values in the corresponding administrative units in the SBR are updated.

> ‣ In the case of an administrative unit being marked as cancelled, or simply not appearing in the incoming administrative data, the activity status of corresponding administrative unit in the SBR is changed to "dead".

Statistical units

7.115. The maintenance procedures for statistical units depend on the relationships associating legal, operational, and administrative units with statistical units. Specification of updating procedures can be expressed in terms of a matrix with:

> ‣ One axis containing all possible types of legal and administrative signals, namely, informative changes, that can be detected in incoming administrative files.

> ‣ The second axis containing all the types of statistical unit change that can occur.

> ‣ The cell entries defining the rules for how each signal is to be reflected as an appropriate update to statistical units, taking into account continuity rules, stability rules, and rules for dealing with conflicting information.

Statistical unit birth, death, and continuity rules

7.116. Maintenance procedures must incorporate continuity rules identifying the types of legal and administrative signals that result in the births, deaths or continuations of statistical units, as discussed earlier. These rules determine the circumstances under which a unit is deemed to be born; is deemed to have died and possibly replaced by the birth of a new unit; and is deemed to continue but possibly in some new form, or under new ownership.

7.117. The rules have to cover every possible type of birth, death or continuation of a unit that can occur, taking into account the source of the signal of a change and the maintenance group to which the unit belongs. In order to be practically applicable, the rules must take into account the ways in which changes can actually be detected.

Stability rules (resistance rules)

7.118. As previously noted, changes to the enterprise characteristics used for sampling must be subjected to stability rules (sometimes called resistance rules) that inhibit short-term changes, thereby ensuring that changes are permanent before updates are made. These rules must cover every possible type of change that can occur and the characteristics involved. They must take into account how the changes can be detected, the types of change (permanent, temporary, seasonal) and the possible impacts on frames and published statistics.

Quality management considerations

7.119.   When the values of a group of characteristics of statistical units are updated it is important to know how downstream statistical processes will be affected. The impacts of changes on survey frames and on the statistics produced by surveys that use those frames have to be considered. In some cases, it may be desirable to hold back updates until a certain point in the annual cycle of surveys so that their impact on the consistency of survey estimates is minimized. For large units, such decisions should be made on a case-by-case basis, involving profilers and users. For small units, rules may be automated.

7.6   Treatment of errors

7.120.   An error may be defined as "a difference in the information presented in the register and the information as it should be, according to a chosen image of the real world produced and maintained by an accepted instrument and documented procedures".[54]

7.121.   It is impossible to avoid errors in the SBR. Indeed, as discussed in connection with stability rules, it may not even be desirable for the SBR to be perfectly up to date at a particular point in time.

7.122.   Changes to statistical units and their characteristics may be either a reflection of real-world events, in which case they are regarded as updates, or an amendment of information that was previously wrong, in which case they are regarded as corrections of errors.

7.123.   This section deals with correction of errors and discusses how they should be applied using established and documented procedures.

7.124.   In essence, there are three steps in handling errors:

> ‣   Determine whether errors have occurred.
> ‣   Decide whether they should be corrected and, if so,
> ‣   Decide how and when to correct them.

7.6.1   Taking account of different user needs

7.125.   The handling of errors should take into consideration the different types of surveys based on the SBR. Procedures appropriate for structural surveys may be different from those appropriate for short-term surveys.

7.126.   Special provisions for the handling of errors in large and economically significant units may need to be negotiated with users, in particular with the managers of economic surveys and the national accounts. Errors in such units may have a considerable impact on statistics. If a special procedure is implemented to take account of the effects on estimates of corrections to these units, it should be fully documented. Special procedures can be applied by a dedicated profiling team or by appointing an authority (such as the SBR manager) who has the final say in making corrections.

7.6.2   Data about errors and corrections

7.127.   As different surveys demand different approaches, it is desirable to record details of each error, including source, type and dates of detection, occurrence and correction. This enables SBR users to access both corrected and uncorrected data according to their needs. In practice, depending on the information and resources available, it may be possible to record only some types of error and some dates, or to record these metadata only for core characteristics.

[54] This section is based on chapter 18 of the *Business Registers Recommendations Manual* (Eurostat, 2010). For this definition, see p. 158.

7.128.   A database may be set up specifically to record the details of errors and associated corrections. The structure and functionality of such a database depend on the procedures agreed upon for handling errors. The database should be closely linked to the SBR or an integral part of it. Ideally the following metadata should be recorded:

> ▸   Type of error
> ▸   Original value (the wrong one)
> ▸   New value (the correct one)
> ▸   Date of occurrence
> ▸   Date of detection
> ▸   Date of correction
> ▸   Source of corrected value
> ▸   Mode of correction (interactive or automatic)
> ▸   Name of person making correction (if interactive)

7.129.   It is also useful to record a date of confirmation for characteristic values that were thought to be in error because they looked unusual or out of date, but that have been investigated and confirmed as correct.

7.130.   Although the inclusion of all these metadata makes SBR maintenance more complex and results in even more opportunities to make errors, the benefits outweigh the disadvantages. The information enables re-creation of the SBR at any given past moment in time, thus allowing users to ignore corrections if they need to do so.

7.6.3   Time dimension

7.131.   The timing of corrections is important. Corrections of errors in characteristics not subject to stability rules, in particular identification characteristics such as names, addresses and telephone numbers, should normally be carried out immediately. If not, data collection may be hindered and respondents upset because information they have provided is apparently not being taken into account.

7.132.   Corrections of errors in economic stratification characteristics such as economic activity and size are a different matter. If these corrections are made the moment the errors are detected, they may affect the consistency between statistics with different periodicities and they may be indistinguishable from genuine changes in the economic world. To avoid these potential problems, corrections may be stored until an agreed moment, for example once per year, when allowance can be made for them.

7.133.   While it would be useful to know the moment at which each error occurred, in practice, the date of occurrence is often not known. The date of correction, assuming that a correction is made, may be considerably later than the moment of occurrence. If the dates of occurrence and of correction are in fact recorded, it is possible to construct a population of units for a point in time in the past that is more accurate than was the actual situation in the SBR at that past time point.

7.134.   Another question that may be asked is: "How far back in time should corrections be applied?" The answer depends on the correction strategy, not only for the SBR but also for statistics based on the SBR. If, for example, there is a revision policy that requires all statistics to be revised after a period of five years, then corrections should made as far back as the moment of the previous revision. If such a revision procedure applies only to the national accounts and not to the individual statistical series on which the accounts are based, then the period for which corrections should be carried out depends on the periodicity of the statistical series.

7.6.4   Error correction policy

7.135.   A systematic policy for the handling of errors should take into account the following:

> ▸ There should be an inventory of SBR users and uses and the consequences of the various types of errors for the various groups of users.

> ▸ The SBR should be structured and maintained in such a way that the correction of errors has a minimal impact on statistical surveys.

> ▸ SBR inputs, processes and outputs should be systematically monitored to detect potential errors.

> ▸ SBR processes should be fully documented so that the handling of changes or errors in the values of characteristics is clear to all concerned. This helps in detecting errors and avoiding discussions on the quality of individual records.

> ▸ The policies regarding handling of errors in the SBR should also be fully documented and audited periodically to make sure that they are still appropriate.

> ▸ Responsibilities regarding handling of errors should be clear and documented. It is advisable to appoint an authority (such as the head of the SBR) who has the final say in difficult cases.

> ▸ The handling of corrected values in statistics based on the SBR should be fully coordinated and documented.

> ▸ The different types of errors detected should be analysed periodically to monitor changes in the pattern of errors over time and thus to inform the future development of policies and procedures for handling errors.

> ▸ Recording the history of errors facilitates the handling of errors in complex situations.

> ▸ If the SBR is used for administrative or commercial purposes, it is advisable to take legal precautions to avoid damage claims arising from errors.

7.136.   The source from which a potential correction originates is an important factor in deciding whether and how to correct the error to which it refers. Care must be taken to ensure that the SBR remains a known reflection of reality and that corrections do not lead to distortions. As previously noted, if the source is a sample survey based on the SBR, there is a danger that corrections of stratification characteristics may lead to bias in future survey samples. Corrections from sources that are not in any way related to surveys can be processed without such a problem.

7.137.   Further discussion of frame errors and their treatment may be found in section 8.4.

Chapter 8
Survey frame methodology

8.1 Introduction

8.1.   The SBR is the backbone for economic surveys, forming a major part of the infrastructure for practical implementation of processes for compiling economic statistics. This chapter presents a framework that describes the underlying concepts in a systematic way, so as to provide the context for elaboration of the survey frame methodology. The starting point is the GSBPM,[55] which clarifies the needs for inputs, processes and outputs and provides a framework for their documentation in terms of eight different phases of a statistical production process, as indicated in figure 8.1.

8.2.   Each phase is subdivided into a number of subprocesses, as shown in figure 8.1. The first phase of GSBPM is to specify the needs and the output objectives and to identify the concepts needed to design a survey. A survey aims to provide information that fulfils the needs and the demands explored in advance with its users. In this phase an inventory of requirements articulated by stakeholders is translated into a framework that enforces coherence in the subsequent statistical processes and provides practical guidance for the implementation phases. The second phase is to design the complete methodology of a survey in terms of the inputs, processing and outputs, including the appropriate statistical unit, the variables to be collected, the data-collection method and the indicators to be produced as output.

8.3.   This chapter describes survey frame methodology in terms of the following subprocesses within these first two phases of the model:

- ›   Identify concepts (1.4)
- ›   Design outputs (2.1)
- ›   Design frame and sample (2.4)
- ›   Design collection (2.3)

8.4.   The other phases and subprocesses are not within the scope of the Guidelines, although it is worth noting that data from the SBR should be available during the third phase, when the data-collection instrument is being built and the production systems are being finalized.

8.2 Identify concepts

8.5.   It is important to use common definitions in all statistical processes that lead to the planned statistical outputs. For example, it may be desirable to use the changes (by comparison with last year) in the values of a short-term indicator (such as quarterly production) as a predictor of the overall change in the corresponding annual indicator. Thus, there needs to be coherence between short-term and annual indicators and hence between the underlying frames.

[55] GSBPM distinguishes eight high-level phases of the process and notes that they are guided by quality and metadata management. See www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0.

A. Let $N(P_1)$, $N(P_2)$ represent specific frame populations of statistical units extracted from the live register for reference periods P_1 and P_2 respectively, with P_1 being before P_2. Let the operator \ between two sets, $A \setminus B$, indicate the set of all elements in A which are not members of B, and the operator ∪ between two sets the union of all elements in the sets. If N entries and N exits represent the number of entries or exits of statistical units, respectively, the following identities hold:

$N_{entries} (P_2) = N(P_2) \setminus N(P_1)$

$N_{exits} (P_1) = N(P_1) \setminus N(P_2)$

$N(P_2) = (N(P_1) \setminus _{Nexits}(P_1)) \cup N_{entries}(P_2)$

B. Let $Y(P_1)$ represent the sum of a variable (such as turnover) based on a frame population $N(P_1)$ for reference period P_1 and $Y(P_2)$ represent the sum of a variable based on a frame population $N(P_2)$ for reference period P_2, then for the variables based on frame population $N(P)$, where $P = P_1 + P_2$:

$Y (P) = Y(P_1) + Y(P_2)$

$Y(P_2) = Y(P_1) + [(Y(P_2) - Y(P_1)) / Y(P_1)] * (Y(P_1))$

8.6.   The above statistical identities are obvious conceptually, but are not usually satisfied in practice. It is quite likely that the sum of the values of a flow variable such as turnover observed monthly (using possibly a sequence of monthly frames) differs from the annual value of the same variable (observed using an annual frame). There are many reasons for such inconsistency, for example, differences between the monthly data and the annual data as regards concepts, reporting periods, sources and frames.

Figure 8.1
GSBPM, version 5.0

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify and code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree on action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalize collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalize outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalize production systems		5.7 Calculate aggregates			
				5.8 Finalize data files			

8.7.   Differences that can be attributed to the frame populations should be controlled and minimized by the SBR. There should be an established relationship between the monthly survey frames and the corresponding annual frame. This is straightforward for a national statistical office that renews its frozen frames only annually; more complex where the national statistical office generates quarterly or monthly frames.

8.8.   To ensure coherence between the outputs of different surveys (e.g. production and employment) for the same reference period, a single frozen frame should be used by all the surveys.

8.3   Design outputs

8.9.   The outputs of a survey may be specified in terms of a population of interest (e.g. manufacturing establishments or enterprises with employees) and a set of variables (e.g. turnover, employment and value added) and a reference period. The survey population is of direct interest to the SBR as it should be derivable from the frozen frames generated from the live register. The choice of variables is also relevant to the SBR insofar as the variables coincide with or relate to characteristics maintained in the SBR. The SBR itself may be a source of statistics, as further discussed in chapter 9.

8.4   Design frame and sample

8.4.1   Choice of statistical unit

8.10.   The target population is the representation of the domain of interest in terms of statistical units in space and time. Statistical units are the building blocks for aggregation of microdata up to the level of industries, institutional sectors or geographical areas. The type of statistical unit that should be used depends on the economic indicators within the domain of interest. The enterprise group is particularly useful for financial analyses and for studying company strategies, but it is too varied in nature and unstable to be adopted as a standard unit for observation and analysis. The enterprise is used for financial data and the KAU is used for production data. Local unit or establishment is used if indicators are required at the regional level. Information on the definitions, derivations and characteristics of statistical units is provided in chapters 4 and 5.

8.4.2   Target populations, frames and coverage

8.11.   The population that would exactly meet user needs is termed the (survey) population of interest. The interpretation of this population in statistical terms is the (survey) target population. The (survey) frame population is the best approximation to the target population that can be provided by the national statistical office. As described in chapter 2, the survey frame is derived as a subset of a frozen frame generated from the live register, using unit selection criterion based on characteristics such as economic activity and size class. The sample is a subset of the survey frame. Figure 8.2 describes the relationships between these various populations of statistical units, indicating overcoverage and undercoverage.

Figure 8.2
Observed, target and frame populations and sample
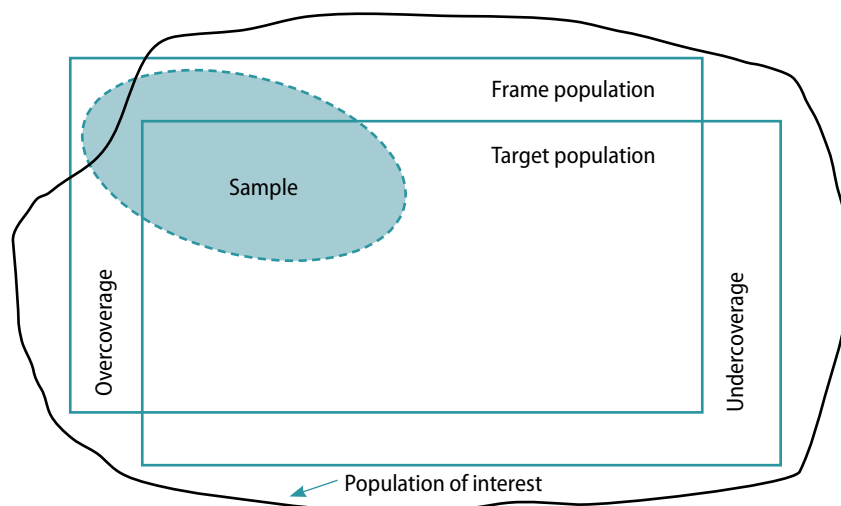
8.4.3   Design frame methodology

8.12.   Figure 8.3 indicates the basic steps and data elements involved in creating a survey frame.

8.13.   The process may be summarized as follows:

> ▸ The live register contains various types of statistical units – enterprise group, enterprise, establishment and others. The units have been derived and are maintained based on information obtained largely from administrative sources.

> ▸ At specified moments in time (month, quarter or year) a snapshot of the live register is extracted containing all units and their characteristics.

> ▸ The frozen frame containing all statistical units that are active, or have been active, within the past 12 months is created by extraction from the snapshot.

> ▸ For each type of statistical unit available in the frozen frame, the corresponding set of units may be referred to as the frame population for that unit type.

> ▸ A survey frame is typically selected from a single frame population and thus consists of a set of one type of statistical units, such as enterprises for a financial survey or establishments for a production survey. Occasionally a survey frame may be designed using two types of unit.

> ▸ It is possible to derive data for each type of statistical unit using different types of observation units (e.g. by primary observation of the statistical unit, or aggregation of information from the tax office linked to legal units).

> ▸ The frames for annual indicators are based on an annual frame population. The frames for monthly (quarterly) surveys of short-term indicators are ideally based on monthly (quarterly) frame populations. Use of monthly (or quarterly) frames and annual frames derived from the same

set of monthly (or quarterly) frozen frames is sometimes referred to as a multiple frame approach.

8.14.   The survey strategy determines at which moment in time a survey frame should be created during the production process. The proposed time frame for delivery of the output indicators should be specified in the business case (e.g. in Europe the regulations on short-term statistics and structural business statistics prescribe the indicators and time frame for their delivery.) Given the delivery date, the indicators and their quality, the production processes and the resources available, the moment in time when the survey frame should be available can be computed.

Figure 8.3
Creation of a survey frame
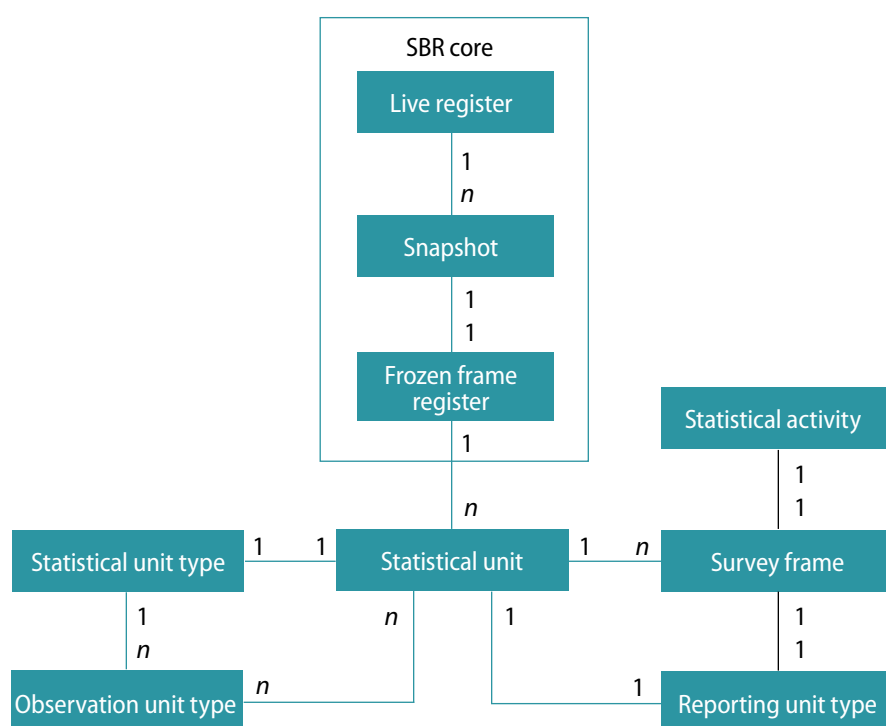
Figure 8.4
Evolution of frame populations
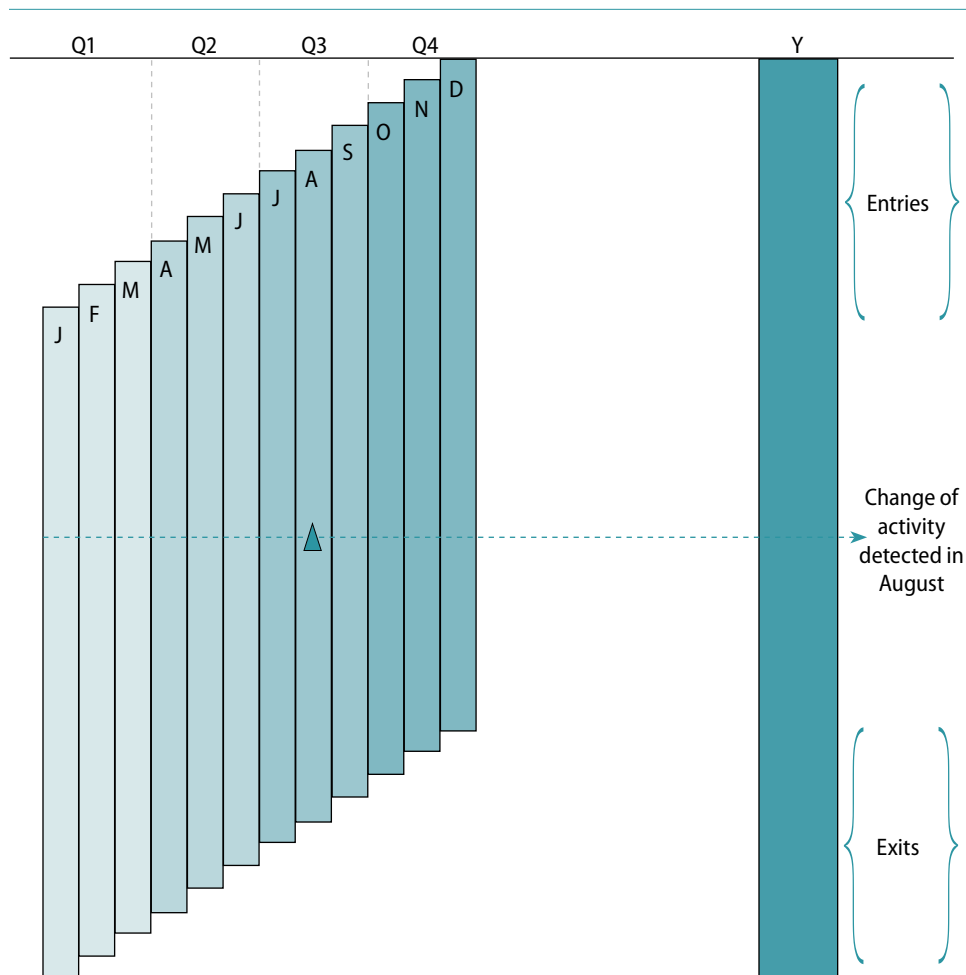
8.15. If the frame population for an annual survey about reference period t is also used to observe the short-term indicators in year t+1, it will not include the information on newly active units for reference year t+1. It may therefore be considered appropriate to create a new version of the frame population at a later point in time for the short-term survey. It depends on the availability of sources whether the live register is maintained in such a way that a frozen frame can be extracted more frequently than annually. Some national statistical offices produce frame populations only annually, some quarterly and some monthly.

8.16. Figure 8.4 illustrates the evolution of monthly frame populations over the course of a year as a result of adding and deactivating units resulting from demographic events.

8.17. The annual population may be defined as all units that are or were an element of one of the monthly frame populations during the reference year. In this respect, it is important to guarantee consistency among the frame populations by appropriate frame error correction procedures (as further discussed below).

8.18. A useful addition to SBR functionality is the possibility of creating improved quality versions of a certain frame population by storing and using a separate list of all frame errors in the live register, including the reference periods to which the errors

refer. Based on such a list it is possible to carry corrected values over to the frame population at the time it is needed.

8.19.　The best way to manage frame errors in a frame population is to avoid their occurrence. This suggests trying to detect anomalies in a frame population in advance of it being used. A good approach is to undertake advance validation based on a set of tables extracted directly from the live register. Units that will appear in the next frame population are compared with the units in the current frame population. Apparent anomalies in the next frame population that could lead to major distortions in the indicators can then be checked (manually) and corrected as appropriate before the new frozen frame is extracted.

8.20.　Given that it is impossible to eliminate all frame errors it is important to identify and analyse the impacts of the various types of frame errors and to use SBR resources in addressing those that are likely to have the more significant impact on survey processing and estimates.

8.4.4　Overcoverage and undercoverage errors and their correction

8.21.　As indicated in figure 8.2 some statistical units in the target population may not be included in the frame population (undercoverage) and some units in the frame population may not be in the target population (overcoverage).

8.22.　Possible reasons for undercoverage are as follows:

- ▸ Sources used to maintain the live register do not include units without employees.
- ▸ Sources used to maintain the live register have size thresholds that exclude smaller unit.
- ▸ Sources used to maintain the live register exclude units which are not obliged to register by law or did not register because of other reasons.
- ▸ Active units which have been accidently marked as inactive in the live register or otherwise excluded from the frozen frame.
- ▸ Recently created units are not yet included in the live register due to the time lag in obtaining and processing information about them.
- ▸ Recently created units are included in the live register but not in the frozen frame from which the survey frame has been derived.

8.23.　Possible reasons for overcoverage are as follows:

- ▸ Inactive units are classified as active in the live register and included in the frozen frame and survey frame. This is a particularly common type of error, as a statistical unit may be linked to an administrative unit that is active for administrative purposes but has never started economic production, or has ceased economic production. In this situation the statistical unit may be marked as active (in accordance with the administrative unit) when it is actually inactive.
- ▸ Some units are duplicates of the same economic entity, which can happen when several administrative sources are used to create statistical units in the live register.

8.24.　Undercoverage is a difficult problem to address because it causes a negative bias in the estimates that cannot be readily estimated from the survey sample. Overcoverage due to the presence of inactive units can be more easily detected during the survey process and allowed for during estimation.

8.25.   Two ways to deal with overcoverage and undercoverage are the following:

> ▸   To accept their existence in advance (which is reasonable provided that they have a minimal effect on the estimation of the indicators); or

> ▸   To reduce them by using information from other sources when deriving the survey frame (although ideally this information should be used to update the live register).

8.26.   If the coverage issues are too large to ignore and cannot be reduced by using other sources of information, then they must be addressed later in the survey process chain by some form of compensation. Known undercoverage should be compensated for by statistical weighting and grossing-up procedures or by extrapolation from census results. Overcoverage may be estimated and corrected by conducting a separate survey with the specific aim of estimating the number of units in the population.

Box 8.1
Country examples to reduce undercoverage: Colombia and Mauritius

Colombia

The SBR team in the National Administrative Department of Statistics is continuously working to increase the coverage of the SBR. When the administrative registers for the different entities, both public and private, are not sufficient, however, there is an option to use phone lines through the call centre available in the entity, when the sources for a survey of a particular subject are required. This practice is being developed by the SBR team to expand the database initially constituted. Coverage is also improved with new sources and with the expansion of the cartographic base through the operation of censuses.

Mauritius

In addition to information collected through surveys, the SBR is heavily dependent upon administrative data from multiple administrative sources for a maximum amount of information. Each source has its own limitations, however, and merging data from the different sources is a tedious process as the unique identifier (the business registration number – BRN) is not yet fully in use across all data sources.

8.4.5   Misclassification by principal economic activity

8.27.   Misclassification of principal activity of units in a frozen frame leads to undercoverage or overcoverage in the subsets of the frame corresponding to specific industries. Special efforts should be made to correct for this kind of error as it leads to increase in variance (due to overcoverage) and bias (due to undercoverage) in the following way. After a frozen frame is produced, a survey frame derived, and a survey sample is drawn, the resulting observation units complete a questionnaire. This may have the following outcomes:

> ▸   It may turn out that a respondent reports that the registered activity code of the observation unit is not correct. In this case, this frame error should be corrected in the survey data set, because otherwise the observation will be attributed to the wrong industry. This does not cause bias but increases the variance of estimates.

> ▸   Alternatively, it may be that units were not included in the survey frame because of being misclassified to out-of-scope industries. This introduces negative bias in the estimates.

8.4.6   Dealing with frame errors

8.28.   Frame errors may cause errors in estimates of levels of indicators, and also, and sometimes more substantially, in estimates of changes in indicators over time. When indicator values are compared for different quarters of the same reference year, or for the same quarters of different reference years, it is important to be able to assume that the changes observed correspond to changes in the real economic world and are not just artefacts resulting from the effects of frame errors or corrections of frame errors. In addition, if a change (by comparison with the previous year) in a short-term indicator (such as quarterly production) is to be used as a predictor for the overall change in the corresponding annual indicator, then it is equally important that the change represents a real change in the population and not the result of an error or correction an error.

8.29.   The way errors in the frame are to be treated should be specified in advance in an official frame error procedure. The aim of a frame error procedure is the following:

- ‣ To support industry-based statistics in providing frame populations that closely approximate the industries in the real economic world.

- ‣ To support coordination of data collection and processing in different industries and in different surveys.

8.30.   Although it is impossible to monitor all the changes that occur in the economic world in real time, it is possible to correct for those frame errors that are likely to cause major distortions in economic indicators, preferably well before publication. These errors may be categorized as follows:

- ‣ Errors that have been present for a rather long time in the live register and are detected more or less by accident, for example during data collection or when a respondent makes contact. These kinds of errors tend to affect the estimates of indicator levels. They do not occur often, as large or otherwise important enterprises are typically included in survey samples and thus updated in the SBR using survey feedback, and, besides, the impact of errors in small enterprises is usually insignificant.

- ‣ New errors that distort the relationships between, monthly, quarterly and annual indicators. These kinds of errors affect estimates of changes in indicators.

8.31.   An extra complication is that, on occasion, correction of a frame error may result in an apparently anomalous change in the value of an indicator, which itself then has to be corrected.

8.4.7   Sample design

8.32.   The sample is selected in accordance with a predefined sample design, which can be quite different from one survey to another. Appropriate stratification characteristics help in selecting a sample that minimizes the sampling error for a particular indicator. All the stratification characteristics required for sampling should be available in the survey frame.

8.33.   Inclusion of inactive units in the sample leads to difficulties in obtaining responses during the observation phase and loss of effective sample size. The sample size should be chosen with this in mind.

8.5   Design collection

8.34.   The data-collection methodology should be designed to optimize survey results in terms of costs, efficiency and accuracy of the indicators. The availability of one or more sources that have the required information influences the data-collection process and the composition of the statistical indicators. It is important to decide in advance whether to collect information directly from statistical units or to use data from linked administrative units. In the latter case, administrative data may be used in place of direct survey data collection or may be used to impute values that are not obtained during a survey.

8.5.1   Observation and reporting units

8.35.   Besides representing the transactors in an economy in terms of statistical units it is useful to introduce the concepts of an observation unit, about which data are obtained, and on the basis of which data about a statistical unit can be derived, and a reporting unit, from which data about the observation unit are obtained. Thus a survey frame consists not only of a set of statistical units, but also corresponding observation units and reporting units together with appropriate contact information.

8.36.   In most cases the statistical unit, the observation unit and the reporting unit are the same, but there are some cases where they may be different:

> ▸   Because of bookkeeping practices no data is available from the target unit, only at another level of the organization.

> ▸   Another organization (e.g. accountancy firms) is responsible for reporting.

> ▸   The partial autonomy of the unit.

> ▸   Data are collected from a group of the units to save costs.

> ▸   There is no frame information about the target unit, only about the unit to which it belongs.

> ▸   The unit of analysis may be a transaction rather than an enterprise, for example freight carried rather than freight carrier.

8.37.   The relationships between these types of units are illustrated in figure 8.5.

8.38.   The creation of different observation and reporting units to obtain data is based on discussion with the respondent about how best to obtain the requested information.

Figure 8.5
Statistical, observation and reporting units
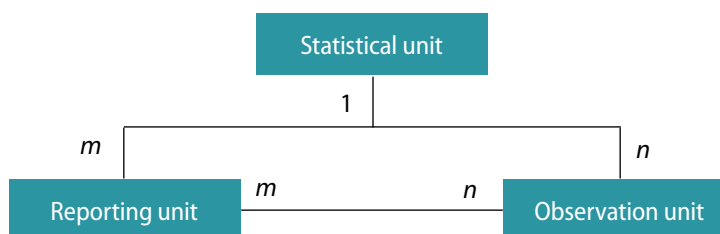
8.5.2 Use of administrative data in place of directly collected data

8.39. The main problems that arise in using administrative data in place of direct data collection as illustrated in figure 8.6 are:

- ▸ The differences between statistical concepts and the concepts defined by administrative rules and laws, which results in the need to identify the correspondences in terms of a metadata translator.

- ▸ The need to transform administrative data into statistical data using an appropriate statistical methodology, in other words, microdata translator.

8.40. An example is the desire to disseminate statistics on turnover. The total amount of turnover could be derived from the administration of VAT revenues. This derivation might, however, lead to serious inconsistencies for very large enterprises because of:

- ▸ The legal organization of the enterprise
- ▸ Consolidation effects
- ▸ Global influences that distinguish them from small and medium enterprises in the way in which they pay taxes

8.41. In this case the metadata translator should make an appropriate adjustment for large enterprises, or the conclusion drawn that such translation is not feasible and that VAT data cannot be used to derive total turnover for large enterprises.

Figure 8.6
Relationship between statistical and administrative universes
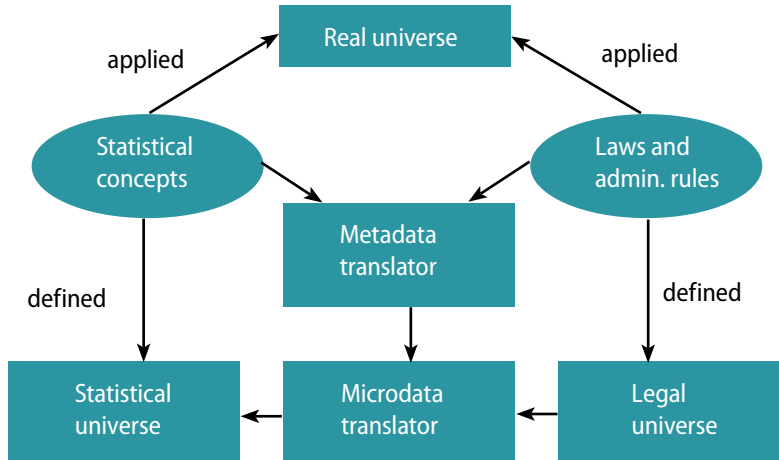
Box 8.2
**Examples of problems with the frame quality: Indonesia, Mauritius and
the State of Palestine**

Indonesia

For countries that are at an earlier stage of developing their SBR, data quality is often an issue, in terms of both coverage and timeliness. If the population in the SBR is incomplete, then the sample is not representative and, as a result, the statistics produced will be biased. If the frame is not up-to-date, the survey will not be able to captures new units or units that have closed down.

Since the SBR is still in the early stage of development, BPS had difficulty in establishing cooperation with the Tax Office before 2019. Regulations prevented BPS from having access to tax data. As a result, the SBR in Indonesia had not been able to use tax data in compiling the framework of economic survey samples. Nonetheless, tax data constitute the best administrative data source because they meet the requirements of administrative data in terms of completeness, coverage and content. Although the SBR can be updated through the survey feedback and quality improvement survey, many units cannot be updated regularly owing to the unavailability of such a good data source on a regular basis. This condition resulted in unsatisfactory data quality in the SBR.

At the beginning of the process of developing the SBR in Indonesia, the SBR was used as the frame for the business tendency survey. Because the SBR data were sourced from the old economic census data and subject matter area data that were rarely updated, many samples could not be found in the field by the enumerators. In the light of this experience, several efforts have been conducted to update the data, including with the involvement of the BPS provincial offices.

In 2018, BPS continued to negotiate the data exchange that had been under way for years. On this occasion, the Tax Office agreed to provide individual taxpayer data, but data on revenue or turnover were excluded. The Tax Office would only provide general data on such characteristics as identity, activity, product, and others because the tax law prohibits

the sharing of revenue data of individual taxpayers. Information on revenue or turnover data was needed to stratify the sample frame. To resolve this problem, the Tax Office would assign a flag variable in the place of the revenue data. The flag variable would describe whether a taxpayer was included in the category of small, medium or large revenue. As a result, the Tax Office will share data on 7.5 million corporate taxpayers with BPS in 2019. The data that will be shared are the data for year n-1 with a frequency of one submission every three months.

The Indonesian Coordinating Investment Board (BKPM) is another alternative data source. BKPM data can be accessed by BPS because there is already a memorandum of understanding between the two entities, but the coverage of the data is not as complete as tax data. The results of the 2016 administrative BKPM data study showed that BKPM data cover only about 100,000 businesses of the following types: limited liability company, limited partnership ("commanditaire vennootschap" or CV in Dutch), firm, personal business, cooperative, foundation, State-owned enterprise and regional-owned enterprise. In terms of business characteristics, BKPM data include business identity but no revenue or turnover.

Box 8.2 (*continued*)

Mauritius

The lack of a functional SBR system or the existence of one with unreliable data (for instance, not reflective of the current status of businesses in the economy) inevitably leads to several problems in the sample frames of surveys, mainly for the following reasons:

(a) Undercoverage: active businesses which are not in the SBR will not be surveyed, which will lead to unreliable survey results;

(b) Inclusion of inactive or moved businesses: survey enumeration effort will be expended to survey businesses which are no longer active or which have changed locations;

(c) Invalid sample representations: if the sample proportions must reflect certain characteristics of the population (such as the proportion of businesses in each sectors, regions and others), the sample will be invalid if the SBR data do not include the latest changes in the characteristics (e.g. operational activity, location and others) of businesses included in the sample frame.

This is a big issue for Statistics Mauritius as, until now, it has not been possible to generate a good frame for the census of economic activities conducted every five years. The issue is mainly for the small establishments (operating with fewer than 10 employees) which are surveyed on a sample basis.

State of Palestine

PCBS strives for an up-to-date SBR in line with international recommendations.

The inspection sampling frame is built on the basis of the establishment census, which is carried out every five years and this results in a large gap in the coverage of units.

If such a sampling frame is used, it has several disadvantages affecting the results of the studies in terms of coverage, comprehensiveness and the accuracy of the extracted indicators, since the use of a random sampling frame negatively affects the process of sample design. In addition, at the statistical level there are certain conditions that must be available in the sample for acceptance (comprehensiveness, non-bias, randomization and homogenization). These conditions may not be available if the sampling frame is random.

These cases are handled by a set of procedures, one of which is a partial updating of the sampling frame in areas with many establishments, and another one is the adjustment of weights based on data collected from household surveys from which the number of establishments can be derived.

Chapter 9
Dissemination

9.1 Introduction

9.1.   This chapter deals with the dissemination of SBR data to the general public. It covers two roles of the SBR as defined in chapter 2, namely, the role of the SBR to produce statistics directly based on the SBR (role 5), and the role of the SBR as an information source providing data on enterprises, their locations and possibly other characteristics (role 6).

9.2.   In the past, SBR data were not usually disseminated directly; rather they contributed indirectly to published statistics that were produced by economic surveys for which the SBR provided the survey frames. The increased use of administrative data and increased computing and storage capacity mean, however, that the SBR is now often a powerful database from which statistical data can be derived for dissemination purposes, or which can be used to supplement data collected by surveys, thereby replacing or reducing the amount of survey data collection. Using the SBR for dissemination purposes does not mean that the other roles of the SBR become less important, in particular the backbone role for statistical surveys. The SBR will not be able to replace statistical surveys totally, as the range of characteristics held in the SBR is quite limited.

9.3.   The SBR may be characterized as comprising a great deal of information of high value. In principle, it covers all enterprises, and related units such as establishments, in the formal sector. This includes government units and non-profit organizations. For all these units it contains the values of basic characteristics such as detailed economic activity classification, regional classification and measures of size in terms of turnover and employment. While this range of characteristics is not as extensive as that which may be obtained by survey, the SBR is nevertheless an invaluable and rich database as it provides coverage that, in the past, could only be achieved by an economic census. In fact, in several countries the SBR is replacing an economic census, or at least supporting by it by providing the frame.

9.4.   In addition to directly publishing SBR data as economic statistics, the second form of dissemination is of business demographic data showing the creation, continuation (survival) and cessation of enterprises or establishments. Such data may be provided by industry, region, legal form, size class and other stratification variables. In this context, the specific unit and continuity definitions used in the SBR are of crucial importance.

9.5.   A further form of dissemination, corresponding to the role of SBR as an information source referred to in chapter 2, is when individual data (also termed unit record level data or microdata) are made available outside the national statistical office. This is possible only in countries where it is lawful to disseminate selected data about individual units to other organizations and, possibly, to the general public. Such a release provision may not include all types of units and almost certainly will not extent to all

characteristics. Typically, it will apply to legal units and the individual data items that can be released may include name, address, legal form, economic activity code and, possibly, size code. Restricted access to the SBR is discussed in section 9.4.

9.6.   Dissemination may be in the form of lists or directories of units, either in print or electronically, online or offline, publicly available, or for the organizations specifically requesting them. Such dissemination may be available free of charge or for a fee.

9.7.   Prior to contemplating this form of dissemination, a national statistical office must establish if there is such an appropriate release provision, and if so, observe it. Otherwise the national statistical office will be in breach of confidentiality requirements.

9.2   Dissemination of economic statistics

9.2.1   Dissemination conditions

9.8.   Statistical information is considered to be a public good. Thus, direct dissemination of economic statistics from the SBR should take place without discrimination regarding type of user and with simultaneous access available to all users. Constant improvements in technology provide increasingly better tools for data communication, and these should be used.

9.9.   Before SBR data are disseminated their quality needs to be assessed. The data should be sufficiently accurate, timely and coherent in that they adequately represent the actual economic structure in the country in those aspects that are disseminated. They should provide good coverage of the economy, industrial structure, regional distribution, size and other characteristics.

9.10.   Given appropriate data quality, one way of disseminating SBR data aggregates is to release regular standard reports, each comprising a set of tables. The reports may be produced annually, quarterly or monthly, depending on data quality, which in turn depends upon SBR creation and maintenance procedures.

9.2.2   Confidentiality considerations

9.11.   As in all statistical data dissemination processes, the need for statistical confidentiality must be taken into account. The general requirement is that no data should be published that would allow a user to determine the value of a characteristic for a specific unit. The usual approach in ensuring confidentiality is to publish only data values that refer to more than a specified number of units, typically three or more units. In a table where a certain data cell refers to fewer than this number of units, the cell value is suppressed and replaced by a symbol indicating that the value is available but confidential. A stronger approach is to apply dominance rules. A cell in a table is confidential, if a specific number of the largest units, for example, the three largest units, in this cell contribute more than a specified proportion, say, 90 per cent.

9.12.   Furthermore, suppression of confidential cells must be accompanied by further cell suppression to prevent the values in the suppressed cells being calculated from the values in other cells by a process referred to as residual disclosure. For example, consider a two-dimensional table including row and column totals and suppose that the one cell value has been suppressed because it is confidential. This value can readily be computed from the other cell values in the same row (or column) and the row (or column) total. Thus, the values of other cells that are not themselves confidential have to be suppressed. This is referred to as complementary suppression. Complementary suppression will have to be applied not only within a single table but also between the

tables to be published so that confidential information cannot be disclosed by combining different tables. There are typically many options for the choice of additional cells to suppress and there are algorithms to guide this choice. For example, the aim may be to choose the data cells with the lowest numbers of units for complementary suppression.[56]

9.13.   In summary, before publishing a set of tables, whether in print or in electronic form, confidential cells need to be identified and the values in these cells suppressed. Then the tables need to be reviewed and further cell values suppressed to prevent residual disclosure. Computer applications that perform such operations are available.

9.14.   The cell suppression method as described above is applied in many countries when publishing their business statistics data. The method is easy to apply and also easy to understand for the data user. Owing, however, to the necessary complementary cell suppression, a considerable information loss can occur as data cells that are not themselves confidential need to be suppressed. There are various other methods to ensure confidentiality, such as perturbation (where cell values are changed at random) or record swapping (where unsafe cell values are switched between units). These methods also result in information losses, however. The record swapping methods may be implemented in a way that aggregated values of the characteristics are restored.

9.3   Dissemination of business demographics statistics

9.3.1   Background

9.15.   Business demographic statistics provide data on the numbers of births and deaths of enterprises or establishments in a specific period, and on the number of enterprises or establishments that were born in a previous period and continued in (in other words, survived to) the specific period. The development of newly born enterprises is usually followed for a period of five years, thus indicating how many enterprises born in t are still alive in $t+1$, $t+2$, .., $t+5$.

9.16.   Business demographic data are usually structured according to activity, legal form, size class, region and various other stratification characteristics. The main descriptive characteristics are employment and turnover. Thus, it can be shown how many jobs were created by the newly born enterprises, and also how many jobs were lost owing to enterprise deaths.

9.17.   Demographic data about enterprises should also be related to the appropriate total numbers of enterprises to produce birth rates or death rates. in addition, survival rates (percentage of enterprises born in t that have survived until $t+1$, $t+2$, and so forth) are useful indicators to calculate for analysis of entrepreneurship performance.

9.18.   Business demographic statistics are sought by analysts and policymakers concerned with business dynamics and entrepreneurship, for example to assess the impact of policy initiatives on enterprise birth and survival rates.

9.19.   Using the SBR to produce business demographics requires:[57]

> ➤   A typology of demographic events that covers all the main demographic events affecting statistical units and takes account of the links between them (see chap. 7.3)

> ➤   Definitions of the main demographic events, specifying precisely their characteristics

> ➤   Methodology and operational rules for computing the statistics

[56] Such algorithms are readily available in software packages, namely the R package sdcTable (Statistics Austria, http://cran.r-project.org/package=sdcTable) and the software tauArgus (CBS, https://research.cbs.nl/casc/tau.htm). For more information on statistical disclosure control, see European Business Statistics Manual, 2021 edition, chapter 17 - Statistical disclosure control (https://ec.europa.eu/eurostat/documents/3859598/12453409/KS-GQ-21-001-EN-N.pdf/f67631e8-c728-e650-d777-de0d9079bf18?t=1614247963543).

[57] United Nations Economic Commission for Europe, *Guidelines on the use of statistical business registers for business demography and entrepreneurship statistics* (2018), www.unece.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20185.pdf.

9.20.   In order to support the production of business demography statistics, it could be the case that infrastructure of the SBR might need to be adapted, to cover specifically the longitudinal aspects and to be able to record and keep track of the units that continue over two or more consecutive reference periods. Basically, an SBR supporting business demography must allow, first, retroactive updates to business characteristics – classification, size and statuses; and, second, easy tracking of longitudinal units through time (continuation or survival). A continuing (surviving) longitudinal unit may be linked to more than one statistical unit in the SBR over the course of its history. Further explanation may be found in chapter 3 of the ECE Guidelines on the use of SBRs for business demography and entrepreneurship statistics (2018).

Box 9.1
Measuring entrepreneurship with the SBR: OECD-Eurostat Entrepreneurship Indicators Programme

Even though the role of entrepreneurship as a driver of economic growth and job creation entered the policy debate some decades ago, there is a lack of sound international evidence on the entrepreneurial phenomenon and its determinants and impacts, nor are there sufficient reliable data for performing comparative analysis. To respond to the need of internationally comparable official statistics, the OECD-Eurostat Entrepreneurship Indicators Programme (EIP) was launched in 2007.

EIP is aimed at the development of policy-relevant and internationally comparable indicators of entrepreneurship and its determinants, in order to support the analysis of entrepreneurship. To that end, the programme has developed a conceptual framework and a methodology for the collection of harmonized entrepreneurship statistics. EIP is characterize by two important features. First, the core set of entrepreneurship performance indicators collected by the programme consists of business demography statistics on the birth, death, survival and growth of enterprises (as distinct from other approaches to entrepreneurship measurement that focus, instead, on data on individuals). Second, the SBR is used to compute business demographic statistics, according to the methodology presented in the *Eurostat-OECD Manual on Business Demography Statistics* (2008) – a manual developed by EIP.

EIP focuses on "employer business demography", where the relevant statistical unit is the enterprise with at least one employee. Employer indicators are found to be more relevant for international comparisons than indicators covering all enterprises, as the latter are more sensitive to the particular coverage of the SBR, for example when employment or turnover thresholds are applied by the main sources used to compile the SBR. (If there is an employment threshold, a basic criterion for comprehensive employer demographic statistics is not satisfied.)

The development of a database of comparable business demographic statistics was much welcomed by policymakers and the research community. The results were also disseminated through the annual publication *OECD Entrepreneurship at a Glance*, which ran from 2011 to 2017.

> **Box 9.2**
> **Businesses demography databases of Eurostat and OECD**
>
> Before 2019, the annual collection of core business demography statistics in European Union followed a common framework[58] based on the recommendations of the *Eurostat – OECD Manual on Business Demography Statistics* (2008), which is also the reference instrument for the production of the above set of business demography statistics in OECD member and partner countries in the context of EIP.
>
> The outcome of these initiatives is that two databases of internationally harmonized statistics on business demography have been created by Eurostat and the OECD, accessible through their respective data portals. These are, for Eurostat Business Demography: http://ec.europa.eu/eurostat/web/structural-business-statistics/entrepreneurship/business-demography, and for the OECD Structural and Demographic Business Statistics: http://dx.doi.org/10.1787/sdbs-data-en.
>
> These databases are widely consulted by the research and policy communities and are a valuable reference for countries not yet compiling business demography statistics.
>
> Further information may be found at https://ec.europa.eu/eurostat/web/business-demography/database
>
> *Source:* United Nations Economic Commission for Europe, *Guidelines on the Use of Statistical Business Registers for Business Demography and Entrepreneurship Statistics* (New York and Geneva, 2018)

[58] European Union, Regulation (EC) No. 295/2008 of the European Parliament and of the Council of 11 March 2008 concerning structural business statistics and European Union, Commission Regulation (EU) No. 446/2014 of 2 May 2014 amending Regulation (EC) No. 295/2008.

9.3.2   Advantages of using the SBR

Coverage

9.21.   In principle, business demographic statistics may be compiled from sources other than the SBR, in particular business censuses and surveys. There are limitations, however, that render these other approaches less convenient or more expensive. With regard to censuses, the coverage of units is typically very comprehensive, but there is a major problem with census data periodicity (generally every five years or less frequently). The infrequency of these data is a source of comparability issues.

9.22.   While survey-based approaches to the production of business demographic statistics may be useful, for example to capture the creation of informal enterprises, they suffer from the usual constraints of survey errors and of sample sizes, which limit detailed data breakdowns. It is also difficult to identify enterprise deaths by means of surveys, as dead units are not available to be surveyed. Moreover, in calculating business demographic rates that require an estimate of the total population of enterprises as the denominator, conceptual consistency between the denominator and numerator populations can be realistically achieved only by using the same source of information for both. Such demographic rates could be, for instance, the birth rate, which is the share of the total births in a given period in the total number of enterprises in the same period.

9.23.   Lastly, using data from an SBR is generally quicker and cheaper than conducting a survey and imposes no additional burden on businesses.

International comparability

9.24.   Thanks to progressive harmonization of the basic requirements of SBRs across countries, the SBR constitutes the most convenient source of business demograph-

ics from the perspective of international comparisons. It is important, however, to acknowledge some potential limitations in this respect. Specifically, the various thresholds used in SBRs are perhaps the most important source of incomparability in business demographic statistics. Although SBRs aim for comprehensive coverage, in practice, they each have to use some sort of threshold that excludes small or difficult to find enterprises. Typically, thresholds are based on monetary values, using turnover, for example, as the indicator, or they are based on employment levels, or on registration with one or more administrative sources. They may be dependent on other criteria, for example, they may exclude agriculture or unincorporated enterprises. These variations and differences across countries result in a lack of international comparability. One of the aims of these Guidelines is to reduce the likelihood of lack of comparability.

9.25.   An additional source of incomparability is that the appearance of a business in administrative registers, and hence the appearance of an enterprise in the SBR, does not necessarily coincide with the date when the business became active. Furthermore, this can vary by country. For example, in some countries, a business may be required to register, or may voluntarily register, with an administrative source before any production occurs or turnover is recorded. In fact, registration does not necessarily mean that the businesses will ever be economically active. It may instead remain permanently inactive. In other countries, administrative sources register a business only after it has already been active for a while, and perhaps only after its production exceeds a certain threshold, commonly based on turnover or employment.

9.4   Microdata dissemination

9.26.   Access to individual unit data in the SBR is restricted by confidentiality requirements. In recent years, however, some countries have been able to give access to such microdata to academic institutions, researchers and students, but always under a number of conditions to ensure appropriate confidentiality.

9.27.   For some research purposes the SBR data alone may not be sufficient and other individual data may be linked to the SBR data, depending on the research goals. Here again, an adequate legal basis must be assured.

9.28.   Restricted access to individual unit data is not the same thing as the publication of these data. As previously noted, however, in some countries publication of such data from the SBR is allowed by law. The content of such a publication is typically restricted to identification code, name, address and legal form of the enterprise, economic activity code, and possibly other classification codes. Usually, no quantitative economic data, such as employment or turnover, are included. The coverage of individual units may also be restricted to companies and government units or even to bigger units only. The intervals at which such data are updated may vary from daily to annually. Usually, such data are published online.

9.29.   There are countries that provide public access to individual unit identification and location data because this information has been disseminated by the units themselves as part of their contact and promotional information. These data may be used for public and private policy planning, and also for economic research. They typically include identification characteristics, stratification by size, activity class and geographical location in the national territory. Typically, they are made available under the general principle that statistical information should be considered a public good and should be made available to everyone in a simultaneous way and without detriment to the protection of personal data.

9.30.   The principle of access to administrative data is that, while national statistical offices have access to individual data, the individual data flow follows a one-way prin-

ciple whereby individual data cannot be provided by the national statistical offices to administrative data providers. There are countries, however, where the dissemination and sharing of microdata from the SBR with non-statistical users are not only allowed, but even legally required (see sect. 6.4).

Box 9.3
National Directory of Economic Units in Mexico

Although its main function is to serve as a sampling frame, the National Directory of Economic Units (DENUE) has made individual statistical unit data available to all, as a public good, and not only for decision makers at government or political level. With the benefit of modern technology, DENUE provides readily accessible information on the distribution of the economic activities and establishments by territory throughout the country.

DENUE is available to any user through a free consultation system in INEGI and it includes legal name, complete address, geographical coordinates, industry classification and size of the establishment according to employment strata through 22 variables which were approved for dissemination.

In addition to the data that are openly shared, DENUE contains data for internal working purposes only, for example, to identify records of the establishments in more than one database or to solve errors among the databases used.

See www.inegi.org.mx/app/mapa/denue/default.aspx.

Box 9.4
Examples of dissemination policies and practice of business statistics data based on the SBR: Colombia, Denmark, Indonesia and Mexico

Colombia

According to Act 79 of 1993, the anonymity of microdata must be guaranteed in order to protect the fiscal or judicial data of companies. Accordingly, the dissemination of SBR information is carried out in compliance with the Act, publishing only the identification, location and economic activity data of the companies through the link https://geoportal. dane.gov.co/geovisores/economia/directorio-estadistico-de-empresas/.

Pursuant to the Act, internal users of the SBR or users within the national statistical system who require access to microdata are provided with variables in ranges of value or geographical aggregated data that do not reveal the identification of the source.

On the other hand, considering the needs of statistical operations, summary tables are generated indicating the result of the updating process: number of new companies disaggregated by activities that are of interest and that meet the inclusion criteria for the surveys.

Denmark

In Denmark, the basic data on business units (legal number, name, address, activity code and other features) are jointly organized and maintained in a publicly accessible business register. In order to improve efficiency and to reduce the burden on enterprises, public authorities are required to use this common register as the source for their inventories of contact information on businesses and they are also required to use the legal number when contacting the enterprises. Furthermore, the enterprises themselves are required to use the legal number on their website and in their communication and transactions with other enterprises, such as on invoices. In addition to the benefits for the statistics production, these requirements have been important drivers for e-commerce and e-government in Denmark.

Box 9.4 (*continued*)

Indonesia

Based on Indonesian regulations, BPS can release business data in three forms to the public:

- Aggregated data
- Directory data, consisting of information as might be found in the telephone directory, such as business name, address and contact, and also economic activity
- Microdata, which are individual data that include statistical variables (such as income and expenditure) without any indication of the business name and address

The published directory data and microdata cannot be linked, owing to data confidentiality. Data published on the website can only be downloaded in PDF format with a national statistical office watermark. At the time of writing, business data based on the SBR are not yet disseminated to the public. When they are ready to be published in the future, however, the publication will follow the previously mentioned policy.

Mexico

INEGI publishes some monthly indicators for monitoring the short-term evolution of the Mexican economy. One of the indicators is the Monthly Manufacturing Activity Level Index in Mexico (IMAI 31–33 from the North American Classification System – NAICS). The measurement of the IMAI 31-33 index, using the national accounting methods, is an elaborated process that generates a result about 40 days after the end of the reference month.

In order to provide more timely information, INEGI is working with some administrative registers, with the purpose of linking establishments from the electricity consumption data register to the economic units which belong to the SBR of Mexico.

The SBR provides the industrial classification for each one of the linked economic units, while the information on the electric energy consumption is provided by the Federal Electricity Commission, approximately 15 days after the end of the reference month.

Through a statistical analysis, a high linear correlation is observed between the electric energy consumption and the manufacturing activity level. Accordingly, INEGI has developed a linear regression model, considering the index of electric energy consumption (ICEE) as an independent variable which is generated from the Commission's data linked to the Mexican SBR for the most relevant manufacturing establishments according to their revenues. The response variable for this model is IMAI 31–33.

Since August 2015, INEGI has been producing timely estimates for the IMAI 31–33 index, comparing them with their corresponding published values to evaluate their accuracy empirically by using this methodology. Until July 2018, in 94 per cent of the analysed periods, the IMAI 31–33 published index values fell within the 95 per cent confidence estimation limits from the regression model. This means that, so far, the observed empirical accuracy approaches the theoretical confidence level.

These IMAI 31–33 early estimates are communicated to the heads of certain Mexican government agencies, like the central bank, the Ministry of Economic Affairs and the Ministry of Finance, among others. It is noted, however, that these results come from a methodology which is still in its experimental stage. Currently, INEGI is working to publish these results on its website, under the experimental statistics category.

Chapter 10
Quality

10.1   Introduction

10.1.   Quality may be defined as the degree to which a set of inherent characteristics fulfils requirements,[59] and may be thought of in general terms as fitness for purpose. The SBR purposes are to provide:

> ▸   Information that enables identification of statistical units
>
> ▸   Populations of statistical units for survey frames and sampling
>
> ▸   Statistics on the structure of the economy – in terms of units and their economic classifications
>
> ▸   Data for business demography analysis
>
> ▸   Tools for using administrative data for statistical purposes

10.2.   It achieves these purposes through the eight roles described in chapter 2.

10.3.   The standard quality dimensions used in assessing a survey can be used in assessing the SBR. For example, users want SBR data to be relevant, accurate and timely. "Relevant" means that the data meet current and potential needs of the users in terms of units and characteristics to support the production of statistics. "Accurate" means that the information recorded corresponds to the reality. "Timely" means that the data provide a picture of the real world with the least possible time lag. There are significant differences, however, between assessing the quality of a survey and assessing the quality of the SBR, owing to the wide range of inputs to the SBR, the multiple types of units that it contains, the fact that the most important outputs are microdata, not statistical aggregates, and the importance of links to units in other databases. Thus, it is not possible to use exactly the same approach for measuring SBR quality as for measuring survey quality.

10.4.   SBR data quality is closely related to how the data are used and to whether they satisfy the users' needs. The SBR has several different groups of users, however, and each group has its own needs. Where frozen frames are concerned, the requirements relating to timing and accuracy are different depending upon whether the frame is to be used for a short-term survey or a structural business survey. For example, if value added is being estimated by a survey based on an SBR frame or using administrative data, the accuracy of the values of key characteristics such as economic activity code and size code for large units will be paramount, given the economic significance of these units. On the other hand, if business demographic indicators are based on the SBR, the quality of the smaller units will be very important, owing to their high rate of involvement in demographic events. In summary, as the SBR is complex and there are numerous and different groups of users of its products, the criteria for evaluating its quality are also complex. Thus the level of quality depends on the specific SBR subpopulation and the strategy for its maintenance (e.g. by activity, size, maintenance group, profiling).

[59] International Organization for Standardization, ISO 9000: 2005 para. 3.1.1.

10.5.   The differing requests from different groups of users result in conflicting demands regarding the timing with which the SBR record changes and delivers frames. The solution, as already described, is to maintain a live register that reflects the latest available information and to produce a set of frozen frames each pertaining to a particular reference period. The frozen frames serve for sampling and coordinating the results and for business demographics. They can be extracted as frequently as needed (monthly, quarterly) in order to support surveys according to their timetables, and include corrections of previous mistakes, new units and recent updates to key characteristics. The live register is continuously updated and can be accessed for the very latest information when required.

10.2   SBR quality dimensions

10.6.   Quality dimensions (sometimes called quality components) are the concepts used to describe some part or facet of the overall concept of quality, when applied to statistical outputs. The United Nations National Quality Assurance Framework identifies the following quality dimensions: relevance, accuracy, reliability, timeliness, punctuality, accessibility, clarity, coherence and comparability.[60]

[60] *United Nations National Quality Assurance Frameworks Manual for Official Statistics*, https://unstats.un.org/unsd/methodology/dataquality/.

10.7.   Although articulated primarily with censuses, surveys and administrative collections in mind, this set of dimensions applies equally well to an SBR. The need for an SBR to be relevant, accurate and timely has already been discussed.

10.8.   As regards punctuality, SBR users certainly need to know when to expect the frozen frames and statistics based on the SBR.

10.9.   As regards accessibility, the possibility of internal users obtaining individual data by directly connecting to the database, and external users obtaining aggregate tables directly from the national statistical office output database should be considered.

10.10.   In terms of comparability, there are two aspects to consider: region and time. Comparability requires that the concepts and methods underlying the maintenance of the units and their characteristics do not change across regions or over time.

10.11.   Coherence includes both internal coherence of units and characteristics within the SBR and coherence with other registers. While internal coherence involves a consistent treatment of the SBR data, coherence with other registers is promoted by creating and maintaining links. The use of a common identification code across all official business registers (administrative and statistical) is an excellent way to obtain greater coherence.

10.12.   Lastly, even though it is not a quality criterion, consideration of cost must accompany measurement of quality. Cost is a quality constraint and must be considered when allocating resources to improve any quality dimension.

10.13.   Cost includes burden on enterprises, and also cost to the national statistical office. It is important to ensure that enterprises are not routinely obliged to provide the same data to the SBR as they have already given to an administrative source. This does not preclude the possibility of conducting an SBR quality improvement survey to check or improve the characteristics of SBR units.

10.14.   Although described separately, the various quality dimensions are interrelated in the sense that attempts to improve the SBR in respect of one dimension may lead to deterioration in respect of another.

10.15.   Quality measurements involve the availability of metadata necessary to correctly understand the information.

10.3   Difference between survey quality and SBR quality

10.16.   In the context of a national statistical office, quality assurance may be defined as all the planned and systematic activities implemented that can be demonstrated to provide confidence that the processes will fulfil the requirements for the statistical output.[61] Quality assurance procedures have been intensely studied and developed in relation to the conduct of surveys. While many of the resulting concepts and methods may be applicable also to the SBR, the SBR has certain specific characteristics that distinguish it from standard surveys and that require different quality assurance procedures. One example is the heterogeneity of SBR users. Another is that the primary SBR output is individual unit-level data, not statistical aggregates. The following paragraphs describe some of these specific characteristics and their impact on the quality issues.

[61] *United Nations National Quality Assurance Frameworks Manual for Official Statistics,* glossary.

10.3.1   Use of administrative data

10.17.   The use of administrative data for statistical purposes has increased over the last decades. The SBR depends extensively on information that is available from administrative sources. The quality of the SBR, therefore, is strongly linked to the quality of the administrative data that it uses, but over the generation of which it may have limited or no control. This is different from a survey process that is entirely under national statistical office control. For example, the SBR has little or no influence on the definitions of the characteristics that an administrative source provides, whereas it determines the definitions of characteristics that it collects through the SBR quality improvement survey.

10.3.2   Heterogeneity of inputs

10.18.   The construction and maintenance of an SBR require data from many sources (administrative and statistical) to be integrated. Each source provides only partial information with regard to the units and their characteristics, so no single source can meet all needs. Subpopulations of units are covered by different sources. For example, data about units in the agricultural sector may be derived from a source quite different from those that provide data about units in manufacturing or services. The values of different characteristics may be acquired from different sources, for example, turnover from VAT declarations, employment from the social security register. Information on the large and complex units is typically obtained by surveys or profiling, whereas information on small units is obtained using (often exclusively) administrative data.

10.19.   Thus, an approach to quality that is appropriate for a survey based essentially on data collection from a sample is not sufficient in assessing SBR quality. Rather, the approach must cover the treatment and quality evaluation of data from a variety of sources. In this context, quality evaluation of the SBR as a whole may be difficult. In the first instance, it is necessary to split SBR data into subsets (e.g. according to source, unit type or group of characteristics) and to develop a set of quality indicators applicable for each subset.

10.3.3   Variability of inputs over time

10.20.   In the case of a repeating survey, the process and the data items collected do not usually vary much over time, and when there are changes to the process, these should be controlled and monitored. Many of the problems that arise when using administrative data can be related to changes in processing procedures that are made by the administrative source but without the knowledge of the SBR. There may be changes in

registration and cancellation rules, in the units involved and/or their classifications, in the data collected, and in the control processes. Accordingly, an important objective of the SBR is to verify the stability of the administrative process and data, and thus to avoid having changes that are purely administrative causing structural changes in SBR data that do not reflect changes in the real economic world.

10.3.4   Technological requirements

10.21.   The processes for construction and maintenance of an SBR are characterized by:

> ▸　　　Fairly large amounts of data from many different sources
>
> ▸　　　Changes over time in administrative source content
>
> ▸　　　Complex procedures for data integration, processing and production of outputs
>
> ▸　　　Changes over time in classifications and output data requirements

10.22.   Thus, SBR procedures and systems must be adaptable over time. The control and evaluation of the technological aspects of SBR processes are central tasks in SBR quality assurance and demand skilled IT resources, which are often in short supply.

10.23.   Given that all national statistical offices require much the same SBR fundamental components, the scope for industrialization of SBRs based on international best practice is an important consideration. In particular, development of a generic SBR system that could be readily adapted to the particular circumstances in a country might prove very useful for national statistical offices with limited resources to undertake SBR development.

10.3.5   Primary output in the form of microdata

10.24.   The main objective of the SBR is to produce individual data – data about individual units – for use by surveys in constructing survey frames and samples. Whereas with statistical aggregates random errors tend to average out, with microdata there is no such notion. Overcoverage and undercoverage errors do not balance out. The effects of both must be evaluated.

10.3.6   Heterogeneity of users

10.25.   The quality policy adopted by the SBR must take into account the wide range of users' needs. Given that the primary SBR role is producing sampling frames for surveys, the SBR update processes must be as timely as possible, must guarantee a stable reference population and must provide the most complete coverage possible. For surveys producing macroeconomic variables such as value added, the coverage of large units and the accuracy of characteristics, such as economic activity code and size, are critical. On the other hand, coverage of smaller units is very important in producing business demographic indicators.

10.26.   Another important aspect of quality is coherence. Whereas individual surveys might optimize satisfaction of their external user needs by using their own specially defined units and characteristics, this is not the best approach. It is far more important to ensure that all surveys use the same conceptual and physical set of frozen frames generated by the SBR as this leads to more coherent and cohesive economic statistics. It is much easier for national accounts staff to reconcile statistics coming from the various economic survey programmes when all the surveys start from the same base.

10.27.   Whereas for internal users accuracy at micro level is important (for example using a wrong address for a unit may result in legal repercussions), for external users accuracy and comparability of SBR-derived aggregates are more important. This illustrates the need for trade-offs in setting SBR quality targets and allocating SBR maintenance resources.

10.3.7   Continuous updating of SBR data

10.28.   Updating a unit in an SBR requires deciding how to identify and treat the actual changes in the characteristics that have occurred. Such changes can involve the unit's existence, its characteristics and the links between units. Events should be referenced to a specific time point or period. During a given time period, say (*t*, *t+1*), the incoming and outgoing flows of units to the SBR and to the survey frames generated from it are determined, first, by births and deaths of enterprises, and, second, by changes in the classification characteristics.

10.29.   The updating of the SBR over the period (*t*, *t+1*) is determined not only by actual economic changes occurring in the real world, but also by correction and updating of characteristics of the units following from SBR maintenance processes. Thus, SBR data cannot be considered in the same way as survey data. A survey does not acquire any additional information once it is concluded, while the SBR may acquire data during the period (*t*, *t+1*) that refer to a time even earlier than (*t*).

10.30.   More specifically, in the SBR it is possible to correct retroactively errors in the values at time (*t*) of characteristics such as economic activity code, measure of size, or date at which a cessation or a birth is registered. Changes are virtually always recorded at a date later than that on which they actually occurred. For example, there can be a delay in recording births and deaths or in recording changes in characteristics in the administrative registers used for updating the SBR. Effective and timely updating of an SBR depends to a large extent on how the administrative sources register the information that they receive from the enterprises. There may be errors and delays, especially in connection with small enterprises. The efforts made to reduce or correct these errors and delays should be commensurate with their likely impact on survey frames, processing and estimates.

10.31.   Continuous updating of the SBR means that there is a possibility of revising any data set produced by the SBR that is stratified by activity status, or other key characteristics such as economic activity code or number of employees, taking into account these two different kinds of changes – actual changes and error corrections. Distinguishing real economic changes from error corrections and delayed updates is a fundamental task in assessing and improving SBR quality.

10.4   Assessing the quality of administrative data for the SBR

10.32.   SBR quality assessment involves taking into account the entire process of acquisition, loading and processing of administrative data. Administrative sources can be assessed in terms of the usual standard quality dimensions. Alternatively – or in addition – the quality framework developed for registers by Daas and others (2009)[62] may prove useful. The quality dimensions are expressed in terms of hyperdimensions (Karr and others, 2005),[63] which are source, metadata and data. Each hyperdimension is composed of several dimensions of quality and each dimension contains a number of quality indicators. The project on the use of administrative and accounts data for business statistics under the European Statistical System also provides further information and examples relating to quality assurance may be found in annex D.

[62] Checklist for the Quality Evaluation of Administrative Data Sources (2009), Piet Daas and others, Discussion paper (09042), CBS. More details are provided in annex D3.

[63] Alan Karr, Ashish Sanil and David Banks, "Data quality: A statistical perspective". *Statistical Methodology,* vol. 3, pp. 137-173.

10.33.   The "source hyperdimension" measures the extent to which information contained in a data source is exploited. The key associated quality dimensions concern the frequency of delivery (yearly, monthly, continuously), the relevance with regard to the information needed, the extent to which the information actually meets SBR needs, the relationship with the administrative authorities and the procedures taken to mitigate the risk of dependence on the source.

10.34.   The "metadata hyperdimension" focuses on the conceptual and process-related quality components of the source metadata. Prior to use, it is essential that the SBR fully understands the metadata and their quality because any misunderstanding is likely to have a serious impact on quality of output based on data from the source. Metadata include the administrative regulations and the clarity with which changes in legal environment are described. The regulations determine the administrative units, the definitions of the data that they are to provide, the reference time periods and the forms used in data acquisition. Comparability is adversely affected when administrative data for specific reference periods cannot be readily transformed to the reference periods or points required for SBR purposes (for example weekly data, or averages having to be transformed to point-in-time values). The identification numbering system is also important. If different systems are used by different sources, combining data from the sources is more difficult and error prone.

10.35.   The "data hyperdimension" focuses on quality indicators that can describe, in a quantitative or qualitative manner, the quality of data that are input to the SBR. They refer to technical checks, accuracy, completeness, timeliness and integrability.

10.36.   Source and metadata quality assessment is usually done by assigning scores to different key dimensions according to the sources used and related metadata.

Box 10.1
Examples of measures to assess and improve the quality of administrative data: China, Colombia, Egypt and Indonesia

China

In the fourth national economic census, in 2018, the National Economic Census Office coordinated the work of the National Bureau of Statistics of China and other government agencies that could provide administrative data. Prior to the census, the Office received a total of 260 million records from 44 administrative sources, such as the administrative records on market supervision and regulation, on the public sector and on taxation. After cross-checking and data validation, the Office was able to build a sample frame containing nearly 100 million entities, including businesses and establishments in the secondary and tertiary sectors and self-employed households. The sample frame was then distributed to all the census interviewers for further verification in the field. The interviewers were required to check all the entities in the sample frame that fell under their jurisdiction, and to update the characteristics of these entities such as their status, principal economic activity, data reporting channel, and others. After field validation, the sample frame was then used to select samples in the census. In conclusion, the administrative data from various sources proved valuable in reducing the workload of the interviewers and in improving the coverage of the census. In December 2019, the National Bureau of Statistics updated its business register with data from the fourth national economic census.

In compliance with regulations, each department has a large amount of administrative records and extensive research is needed to verify and make full use of these resources. The activities that the statistical department carries out to analyse the administrative sources include, first, studying the system framework for the administrative records; second,

> **Box 10.1** (*continued*)
>
> identifying the scope of registration of the administrative record system (including related indicators, definitions, classification, and other areas); third, establishing an effective link between various administrative records and between administrative records and statistical surveys; and, fourth, laying the foundation for the statistical application of administrative records.
>
> **Colombia**
>
> The National Administrative Department of Statistics has published a guide to document the methodology of statistical operations based on administrative registers, which sets out guidelines for the documentation of statistical operations based on administrative registers under standardized and harmonized criteria on their structure, presentation and content.
>
> In addition, the Department has created a guide for the elaboration of metadata of administrative registers, laying out guidelines for the construction of the metadata that characterize the registers in order to promote the use, comparability and integration of the registers. See www.dane.gov.co/files/sen/lineamientos/DSO_020_LIN_04.pdf.
>
> **Egypt**
>
> CAPMAS measures the quality of the data of the Social Security Authority framework by making field visits while conducting the economic census and comparing the two results.
>
> **Indonesia**
>
> BPS conducts the assessment for the quality of administrative sources for SBR in various steps.
>
> In the selection of administrative data, BPS conducts administrative data studies by Internet searches or direct visits to the potential administrative data sources. BPS also prepares a list of information that must be explored (which may be a list of questions) in the administrative data studies to determine if that particular administrative data source meets the criteria of coverage, content and access. For coverage criteria, a primary administrative data source must have the broadest coverage of active units that are needed for the SBR. For content criteria, the sources must have sufficient information on active status, economic activity, size for sampling purposes and contact information. Lastly, for access criteria, in theory it is legally possible for BPS to obtain the data from the administrative sources.
>
> When the data are available from the sources (administrative sources or economic census), BPS conducts an assessment comprising the following checks:
>
> › Checking the concepts, standardization and unit measurements in order to determine whether they are the same in the SBR or contain adjustments
>
> › Checking the null values and data discrepancies to determine whether they need to be imputed or not
>
> Checking master data and their updated dimensions where data will become inconsistent if a change in master data is not followed by the adjustment of the detail data

10.5   Frame errors and their implications for surveys[64]

10.5.1   Types of frame errors

10.37.   As noted in section 7.6, an error may be defined as "a difference in the information presented in the register and the information as it should be, according to a chosen image of the real world produced and maintained by an accepted instrument and documented procedures".

[64] Eurostat. *Business Register Recommendations Manual,* chapter 18: "The treatment of errors".

10.38.   The following types of errors may be identified.

Errors in existence

10.39.   This type of error is due to false information regarding the demographic characteristics (date of creation and date of cessation) of a particular unit. There are two categories of existence error:

> Units are recorded as economically active, but are not yet active, or no longer active, in the real world. This results in overcoverage, which may lead to response problems for statistical surveys based on the SBR.

> Units are economically active but are not present in the SBR. This type of error results in undercoverage, which will lead to underestimation in survey outputs.

Errors in identification characteristics

10.40.   Errors in names, addresses, telephone numbers and other details can hamper data collection as a consequence of problems locating and contacting the units. Errors in names and addresses also impede the use of the SBR as a tool to link and coordinate data from different sources.

Errors in stratification characteristics

10.41.   This error type includes errors in legal form, economic activity code, size class (based on number of persons employed, turnover and net assets) and geographical location. These errors may affect the inclusion of units in survey samples and certain strata, and in SBR statistics. They may result in inefficient sampling and sample allocation for surveys based on the SBR, and in inaccurate population estimates derived from the SBR.

10.5.2   Impact of errors on survey frames

10.42.   SBR errors can have a big impact on the processes and results of surveys based on the SBR.

10.43.   Non-response and coverage problems may result in significant non-sampling errors. These two problems are interrelated and can be attributed to SBR errors. For example, it may be impossible to contact a unit included in the frame because the contact characteristics are wrong. Units wrongly included in the frame may be selected for the sample and then found to be out of scope, thus reducing the effective sample size. The impact of frame errors on survey estimates is a measure of the accuracy of the SBR.

10.44.   Frame errors and their impact on overall survey error can be classified according to the following types.

> *Undercoverage*: The frame derived from the SBR does not include all units within scope for the survey. Reasons for undercoverage are: omissions (lags and leakage), errors in activity status (falsely inactive units), and errors in stratification characteristics (units miscoded so as to be out of scope). SBR undercoverage results in underestimation.

> *Overcoverage*: The frame derived from the SBR includes units that are out of scope. Reasons for overcoverage are: duplication, errors in activity status (falsely active units), and errors in stratification characteristics (units

miscoded so as to be in scope). If an out-of-scope unit is not identified as being out of scope, the result is overestimation. If it is thus identified, its exclusion results in a reduction in effective sample size and hence an increase in the sampling error.

▸ *Errors in unit characteristics*: Errors in stratification characteristics such as principal economic activity code, size code, and location cause inefficient sampling and sample allocation. Errors in contact data result in increased non-response and non-response follow-up.

10.6   Classification of metadata

10.45.   Comprehensive and accurate descriptions of the SBR inputs, processes and outputs are a vital aspect of SBR quality assurance. These SBR metadata may be classified using the following schema.

10.6.1   Source of data

10.46.   A source code generally consists of an alphanumeric code that identifies exactly the source used in assigning the value of a characteristic of a unit. For example, it may indicate that the turnover value of a particular enterprise has been taken from VAT data, or that the number of employees' value has been obtained from a particular survey. Every value is assigned a source code.

10.6.2   Procedures used for attribution of unit characteristic values

10.47.   These metadata include the procedures used in resolving conflicting information from different sources.

10.6.3   Production process

10.48.   These metadata include descriptions of the processes for:

▸ Acquisition and processing of administrative and statistical data

▸ SBR profiling and improvement surveys

▸ Production of frozen frames and survey frames

▸ Production of statistics

▸ Integration of administrative registers

10.6.4   Updating and error correction process

10.49.   These metadata include descriptions of updating processes and determining whether changes occurring in a period reflect real-world economic changes or corrections.

10.6.5   Reliability of data

10.50.   These metadata include quality indicators and references to the production sources and processes.

10.6.6   Characteristic value updating history

10.51.   These metadata include the date or period to which the value of a characteristic relates and the date on which the value was last updated in the live register.

10.6.7   Other documentation about sources and methods

10.52.   These metadata may help users to assess the quality of SBR data. Documentation on sources and methods may be disseminated with SBR statistics or be made available through a database (web application) that can be consulted by users.

10.7   Quality assessment methods

10.53.   The following paragraphs outline the methods that can be used to assess SBR quality. Some are generic in the sense that they are applicable to all statistical processes and products; others are specific to the SBR.

10.7.1   User survey

10.54.   Given that quality is defined in terms of user needs, collecting users' views of the SBR is a first step in SBR quality assessment. The aim is to obtain the views of major groups of users regarding each of the quality dimensions. The views of survey staff that use SBR data are particularly important as they are the primary users. Often the assessment results differ from one type of survey to another one. For example, managers of structural surveys may have a different view of SBR relevance to those responsible for short-term statistics. Complete satisfaction of all users is not a realistic goal. In interpreting the results, the relative importance of each of the types of users has to be taken into account.

10.7.2   SBR audit

10.55.   A quality audit may be defined as a systematic, independent and documented process for obtaining quality audit evidence (records, statements of fact or other information that are relevant to the quality audit criteria and verifiable) and evaluating it objectively to determine the extent to which the quality audit criteria (set of policies, procedures or requirements) are fulfilled.[65]

10.56.   Auditing is a "powerful tool … by providing important information"[66] to improve the quality of the SBR.

10.57.   From time to time, an overall audit of the SBR processes and outputs should be undertaken. Since use of external audits can be expensive, it may be desirable to conduct an internal audit. Some national statistical offices have a unit within the organization whose role is auditing, and in this case the unit should undertake the audit. In national statistical offices with no such organizational unit an ad hoc team of auditors should be assembled, including both SBR users and other statisticians with limited (or no) exposure to the SBR.

10.7.3   SBR improvement surveys and quality measurement surveys

10.58.   The primary purpose of SBR improvement surveys is usually to improve the quality of unit characteristics, by detecting and correcting errors and by filling in missing values. Such surveys are also, however, a means of measuring SBR accuracy. Indeed, some may be conducted with the primary goal not of correction but of meas-

[65] *United Nations National Quality Assurance Frameworks Manual for Official Statistics,* para. 4.18. Available at https://unstats.un.org/unsd/methodology/dataquality/references/1902216-UNNQAF-Manual-WEB.pdf.

[66] Eurostat. *Handbook on Data Quality Assessment Methods and Tools* (DatQAM), (Wiesbaden, Germany, 2007). Available at https://unstats.un.org/unsd/dnss/docs-nqaf/Eurostat-HANDBOOK%20ON%20DATA%20QUALITY%20ASSESSMENT%20METHODS%20AND%20TOOLS%20%20I.pdf

urement of the quality of characteristics such as activity status, location, economic activity code and size.

10.59.   Such surveys must be designed carefully because of the difficulty in determining the real values for comparison with the SBR values. Furthermore, to avoid self-referential findings, the survey results should preferably be analysed by statisticians from outside the SBR.

10.7.4   Auditing clerical work

10.60.   Quality audits are a useful tool for monitoring the quality of clerical processing and of automatic updating. They may involve checks on representative samples of clerical update actions (described immediately below) or regular analyses of key characteristics at aggregate level (as described in section 10.7.5), or, preferably, a combination of the two approaches.

10.61.   The aim of auditing clerical processing is to monitor the quality of the clerical input to an SBR. It involves checking samples of clerical updates. The checks should be based on regular, random and representative samples.

10.62.   Clerical audits are normally undertaken by experienced staff, who investigate the work of SBR staff to see if the actions that have been taken comply with the processing guidelines. They may also be carried out by assigning the same update actions to different clerks and then comparing the resulting updates.

10.63.   Clerical error rates can be monitored over time and reduced, for example by improving procedures and by informing and training staff. Regular summary reports should be produced to inform managers and users.

10.64.   The quality auditing function should be closely linked to documentation and training functions, perhaps in a form of quality circle, so that issues are identified, resolved, documented and covered in future staff training. By this means it should be possible to ensure sustained improvement in quality over time.

10.7.5   Macroediting

10.65.   Another class of quality audits involves monitoring changes in the SBR at aggregate level. This should be done on a regular basis and linked to the production and dissemination of the frozen frames.

10.66.   Frequency distributions for key characteristics are compared before and after periods of updating (automatic and manual) to assess the impact of the changes on various subsets of units and to ensure that all changes can be adequately explained. In performing the analyses, priority is given to large units and to units that are particularly relevant for specific surveys or under particular observation from users.

10.7.6   Defining and monitoring quality indicators

10.67.   A system of quality indicators for the SBR is a very important means of assessing quality. The construction and use of quality indicators is discussed in the flowing section.

10.8   SBR quality indicators

10.68.   The starting point in defining a set of quality indicators for the SBR is an appropriate conceptual framework. This is discussed in section 10.8.1. As described

in sections 10.8.1-10.8.3, SBR quality indicators may be divided into three groups corresponding to processing phases, dealing with:

- ‣ Quality of input
- ‣ Quality of processing (or throughput)
- ‣ Quality of output

10.8.1   Conceptual framework for quality indicators

Key factors

10.69.   Key factors in defining indicators of SBR data quality are the following.

10.70.   *Time:* The SBR is constantly evolving, thus quality indicators are characterized by a reference date (t) or period (t-x, t). In presence of a lag between the reference period (t-x) and the assessment period (t), the indicator describes quality at (t-x) measured at (t).

10.71.   *Scope*: A quality indicator is applicable to a particular type of unit (say enterprise or local unit) and, within that type, to a subset of all possible units. The subsetting may be formally defined by a filter presented in the form of a logic formula operating on the relevant set of units and their characteristics, for example, *active enterprises born before (t-x).*

10.72.   *Subpopulation*: A quality indicator must be defined and measured at the level of the subpopulations of interest. Having only global indicators may mask weaknesses in specific subpopulations. It is common to define subpopulations in terms of size (small, medium, big), region and economic sector.

10.73.   *Characteristic*: A quality indicator applies to a particular characteristic of a set of units.

10.74.   *Criteria*: To construct a quality indicator, it is necessary to have criteria for assessing the quality of the characteristic, unit by unit. For each unit, and for each characteristic, it must be possible to assess whether the value is right or wrong (true or false), or where it lies on a scale of quality between 0 and 1.

Quality assessment mechanisms and criteria

10.75.   There are various mechanisms and criteria by which to assess quality as follows.

10.76.   *Use of external information source*: A value of a unit in the SBR may be considered as correct if it is sufficiently close to a reference value from an external source. This is the most commonly used criterion. It focuses on compliance (namely, whether the value of a characteristic in the SBR complies with the value of the same characteristic in an external source). It is a proxy for accuracy (whether the value of a characteristic is correct or not) but is the only possible measure when the real value is not known. The compliance rate (percentage of units for which the characteristic assumes a sufficiently close value) is a practical substitute for reliability rate (percentage of units for which the characteristic is correct) when the latter cannot actually be measured.

10.77.   *SBR improvement survey (quality survey)*: Quality is assessed by comparison with a reference value obtained by an SBR survey. This approach is expensive.

10.78.   *Internal consistency*: A value is deemed correct if it is consistent with the other characteristics of the same unit (e.g. turnover, employees, main activity and legal status). Definition of consistency edits is difficult. Often they are based on plausibility criteria, meaning that most true values satisfy these criteria. Even if a characteristic value passes the consistency checks, there is no guarantee that it is correct.

10.79.   *Temporal consistency*: The quality of a characteristic is defined on the basis of a comparison with its values in previous time periods. The aim is to identify impossible or implausible changes from one period to another.

10.80.   *Quality without witness*: It is possible to make an assessment of quality without having a specific reference value or element of comparison. Examples of the basis for such an assessment are:

> ▸   The date on which the value was most recently checked or updated - with the underlying principle that the more recent, the better.

> ▸   The name of the information source – the likely accuracy of data from a source may be known.

> ▸   The methodology adopted – the appropriateness of the methodology may be known.

10.8.2   Quality of SBR inputs

10.81.   The SBR typically has many input sources, mostly administrative. As the qualities of these sources affect SBR output quality, they are very useful indicators in themselves. From the SBR perspective, though it may be possible to have an influence on the quality of an administrative source, it is not possible to control its quality ex ante. Thus, the quality of each source should be assessed ex post by means of suitable analyses to identify errors in the data supplied and to take these into account in using the data.

10.82.   Quality assessment includes consideration of the accessibility and clarity of administrative data, meaning the ease with which the SBR updating process can access and interpret the data. Good metadata management requires that data format and content are precisely reported, and data are accompanied by metadata, examples and advice.

10.83.   Other simple indicators of the quality of the source include the following:

> ▸   *Time lag*: difference between the date on which the data are supplied and the reference period to which they refer.

> ▸   *Completeness of characteristic*: number of missing values; and number of missing values as a proportion of the total number of values.

10.84.   Many problems that arise from using administrative sources are caused by changes in the source that are not known to the SBR. Such changes may include changes in registration or cancellation rules, in classification criteria, or in the administrative control processes. Every effort should be made to learn about such changes in advance through agreements and regular communications with the administrative authorities responsible for the sources. All such efforts notwithstanding, some unanticipated changes may occur. In order to detect and handle such changes, what may be regarded as "big changes" – meaning changes in data values that are impossible or implausible during a given period (for example a quarter or a year) – should be identified and reviewed to determine the underlying causes. The main objective is to check the stability of the sources and avoid a situation in which changes that are merely administrative produce structural changes in the SBR that do not reflect economic reality.

10.85.   Simple indicators based on the comparison of the values provided in two different years may be formulated along the lines set out below.

Percentage or proportion of variations (per characteristic)

10.86.   For example, a simple indicator is the percentage of units that have changed principal economic activity from year *t* to year *t*+1. This can be computed for the whole population or for a specific subset. Further analytical investigation can be initiated if the value in the reference year exceeds the average level of the indicator over time.

10.87.   Comparison of data from a single source over time is also fundamental in analysing the completeness of enterprise births and deaths. Such analysis is essential in helping to detect and minimize undercoverage and overcoverage. Simple counts of the numbers of births or deaths during a reference period do not of themselves provide much information. Viewed as a time series, more complex and useful indicators may be defined.

Indicator for overcoverage and undercoverage:

10.88.   The effect of update delay associated with a source can be obtained by comparing, where available, two values of a characteristic at different points in time, for example:

> ‣ The number of cessations, N*cess (T)*, occurring in year *t* as reported in data received during a period ending at time *T*

> ‣ The number of cessations, N*cess (T+1)*, occurring in year *t* as reported in data received between time *T* and *T+1*

10.89.   In this way the lag in the registration of the cessation dates in the input source can be estimated using a simple indicator such as *1-Ncess(T+1)/Ncess(T)*. Analysis of this indicator can give an idea how long it is worth waiting to receive cessations before using the data to generate survey frames. The shorter the time, the larger the number of cessations that will not have been received and the longer the time, the smaller the number of cessations that will not have been received, but the later the frames will be generated relative to the survey reference year.

10.8.3   Quality of SBR processes

10.90.   For the purpose of defining quality indicators, SBR processes may be divided into three phases, namely, integration of input data; assignment of values of characteristics; and editing, as described in the following paragraphs.

Phase 1: Integration of input data from administrative sources

10.91.   The purpose of this phase is to integrate administrative data from various sources and to create clusters referring to the same unit (the enterprise). Two different subphases can be distinguished and quality indicators defined for each.

10.92.   *First subphase: analyse within a source.* For each input source, records that pertain to the same legal unit (as identified, for example, by a common taxation identification number) are integrated. Possible quality indicators are the following:

> ‣ Number of duplicates as a proportion of the total number of supplied records. A decrease over time indicates an increase of quality of the source from the perspective of the SBR.

> ‣ Number of new records during a given reference period. This provides a measure of coverage in terms of unit creations. Comparisons over time and with a benchmark can be used to identify possible problems in supply.

10.93.   *Second subphase: link between sources.* Records coming from different sources that pertain to the same legal unit (again, for example, as identified by a common taxa-

tion identification number) are integrated in order to build-up a cluster of records for the same enterprise. The taxation register is, typically, the base used to define the set of legal units and to integrate all the other sources. This phase has the particular aim of identifying the set of administrative data records relating to each individual legal unit. Errors in this phase, such as missing or wrong links, can greatly affect the data produced in the following phases. Possible quality indicators include the following:

    (a)   Number of clusters of records in period (t) linked to legal units;

    (b)   Number of clusters of records in period (t) not linked to legal units;

    (c)   Number of clusters of records in year (t-1) not linked to legal units;

    (d)   Undercoverage indicator: [(a) (c)]/(a).

10.94.   The indicator (d) measures the undercoverage due to time lag in the registration of units in the taxation register. It represents the percentage of units already provided in (t-1) by other administrative registers, not linked in time (t-1) and then linked in the succeeding time (t). These are units that could have been linked and would have been included in the SBR in time (t-1) if the base source (typically the taxation register) had successfully contained such units at that time.

Phase 2: Assignment of values of characteristics

10.95.   The second phase of the SBR production process comprises the assignment of the values of the characteristics of each unit and the identification of active units in year t. The quality of the assignment procedure for each characteristic can be evaluated using indicators based on outputs produced at each step in implementation of the procedure.

10.96.   The characteristic of a unit of which to check the quality first is its activity status. This is because the frozen frames and subsequent survey frames contain only active units (more precisely, active, recently active and potentially active units). Thus, other characteristics need only be checked for these units. Quality indicators for each source may be defined along the following lines:

    ▸   *Active status concordance rate*: number of units for which the unit has an active status in both the source and the SBR, as a proportion of the number of units in common to the source and the SBR.

    ▸   *Inactive status concordance rate*: number of units for which the unit has an inactive status in both the source and the SBR, as a proportion of the number of units in common to the source and the SBR.

    ▸   *Activity status discordance rate*: number of units for which the unit activity status differs between the source and the SBR, as a proportion of the number of units in common to source and SBR.

10.97.   Sources that may be available for determining or checking activity status include the following:

    ▸   Small and medium-size enterprise survey – checks for active and inactive units

    ▸   Foreign trade survey – checks for active units

    ▸   Bankruptcy database – checks for inactive units

Phase 3: Editing procedures

10.98.   The third phase of the SBR production process comprises implementation of the editing and imputation rules used in final identification of the frozen frame for a

particular reference period *t*. In order to measure the quality of the rules, quality indicators may be defined to measure the errors produced by each rule.

10.99.   The quality checking plan generally consists of a set of separate quality project modules, usually executed sequentially. The project modules may be changed in number and composition according to their results over time. Each project module comprises a set of rules having a similar structure and affecting data in a similar way. The main project modules include the following:

- ▸   *Cleaning* – using rules determining the exclusion of some units from further checks.
- ▸   *Deterministic* – using if-then clauses that cause the automatic changes in the values of the relevant characteristics whenever certain conditions occur.
- ▸   *Errors* – using rules that generate error warnings whenever certain conditions occur.

10.100. Since the number of edits may be very high, deterministic rules usually focus on peculiar subsets of units and on specific characteristics (e.g. on the cross-combination of economic activity code and size). The editing process produces error warnings for possible clerical follow-up. There is a need to limit the number of warnings in accordance with the resources available, so that SBR staff can concentrate their efforts on economically significant units rather than smaller ones. This is sometimes referred to as "significance editing".

10.101. The calculation of quality indicators based simply on the number of warnings or errors in a given time period may be misleading. More useful indicators can be built by looking at trends, for example, measuring the increase or reduction in the number of units that fail each type of edit over time.

10.102. A synthesis of quality indicator values can be achieved along the following lines:

- ▸   Variation (in absolute terms and percentages) between *t* and *t+1* of counts of units by type of error
- ▸   Variation (in absolute terms and percentages) between *t* and *t+1* of counts of units for which values of characteristics have been automatically changed by deterministic rules
- ▸   Variation (in absolute terms and percentages) between *t* and *t+1* of counts of units for which a warning is generated and/or that have been manually verified

10.9   Quality policy and improvement

10.9.1   Quality policy

10.103. Elements of a quality policy may include, first, the decision to measure the quality of the SBR; second, the decision to communicate information about the quality of the SBR to all users; and, third, based on the measurements and the feedback from users, the decision to initiate a programme of improving the quality of the SBR.

10.104. As discussed above, the measurement of quality is aligned with three principal groups of activities namely the "input", "processing" and "output" phases. For each of these phases relevant quality dimensions and indicators are identified and defined, based on the specificity of the sources, infrastructure and resources available.

10.105. A by-product of measuring SBR quality is the generation of metadata and paradata that allow a better knowledge of the state, content, structure and processes of

the SBR. Detailed examples from Colombia, Italy and the Kingdom of the Netherlands are provided in annex D.

10.106. An example of communicating quality policy is an SBR quality declaration aimed at all users, internal and external, presenting the quality dimensions and associated indicators. Another example is informing users, in as timely a manner as possible, about significant events related to SBR maintenance, such as the availability of new administrative data, the next frozen frame, changes in important units and changes in classification.

10.107. The decision to implement an SBR quality programme must be based on a careful analysis of costs and benefits. Benefits should be seen primarily in terms of results of improved quality from the perspective of users, in particular statisticians conducting economic surveys and the national accountants that use the resulting outputs. In addition, the SBR quality programme should be viewed within the broader confines of a national statistical office-wide programme to improve economic statistics rather than being seen as end in itself.

10.9.2   General approach to improving quality

10.108. Quality improvement is an iterative process based on, first, construction of a set of quality and performance indicators for SBR inputs, processes and outputs; second, setting quality and performance targets; and, third, defining quality assessment tools. These indicators, targets and tools enable the monitoring of SBR quality as the basis for formulating, reviewing and implementing quality improvements. The results of monitoring quality should be summarized in quality reports that are made available to SBR staff and to users.

10.109. The following subsections set out examples of ways in which the quality of the SBR can be improved. The examples often concern more than one quality dimension as improvements cannot be made entirely independently in each quality dimension. There are trade-offs that need to be considered.

10.9.3   Improving timeliness

10.110. In order to have the most representative picture possible of the population of enterprises, the SBR updating process should be as timely as possible while at the same time providing coverage that is as good as possible. As discussed in chapters 6 and 7, this is achieved by systematically applying updates available from relevant administrative sources, for example data from corporation registration systems, income tax, VAT and social security systems, chambers of commerce and other trade associations. The aim is to detect creation, structural change and cessation of units as quickly as possible. Administrative data such as sales tax remittances, income tax returns and payroll deductions provide clear signals that an enterprise is active. Signals indicating when a business becomes inactive are less numerous and timely.

10.111. Another aspect of improving timeliness is reducing the time required to apply updates so that these changes can be quickly reflected in the survey frames. Often survey managers complain that changes that they have detected through responses to survey questionnaires are not immediately used to update the SBR because staff spend time on processing the changes and deciding on the appropriate updates. One way of reducing processing time is to permit the survey staff themselves (after having received the appropriate training) to perform the updates. Another possibility is to let the businesses themselves update their own information directly through an appropriate external portal.

10.9.4   Improving completeness

10.112. There should be constant investigation of new ways to extend the SBR by collecting information about enterprises from additional sources. For example, it may be possible to link the SBR to registers like an employment database or industry-specific database to obtain additional information on employment size or economic activity.

10.9.5   Improving coverage

10.113. Different strategies must be set up for reducing undercoverage and overcoverage, in particular duplication of units. The risk of duplication depends almost entirely on the particular input sources and the matching procedures used in bringing together and deduplicating data from these sources, as in the following cases:

- ▸ When matching is based on a common identification system there is a risk of misinterpreting the identification codes, for example a fiscal code may be confused with the VAT code.

- ▸ In absence of a common identification, matching depends upon record linkage techniques, for which there are numerous high performance software applications but still a significant risk of missed matches and mismatches.

10.114. Automated procedures must be adapted to the peculiarities of the situation in any given country.

10.9.6   Improving quality reports

10.115. Reporting on quality is a crucial aspect of a quality programme. The general aim should be to deliver short subannual (preferably monthly or quarterly) quality reports to alert users to significant, recent changes in the SBR – changes that affect the frozen frame and hence survey frames and ultimately published economic statistics. Within the report it should be possible to drill down by industry, region, legal status and administrative source.

10.116. Subannual reports should be accompanied by more comprehensive annual quality reports that track smaller changes by month and that show trends over years. They may also report on the results of monitoring particular cases, for example holding companies, foreign controlled transport companies and discrepancies between employment and turnover for big construction enterprise groups operating abroad.

10.9.7   Providing survey support

10.117. To assist with surveys and to monitor the quality of the SBR as seen from a survey perspective, it is vital to provide functionality that:

- ▸ Allows a survey to monitor its population, since every survey needs to understand the significant changes that have occurred between survey cycles and seeks to track births, deaths and changes to large units.

- ▸ Provides a frame on the basis of which efficient sampling schemes can be designed and panels monitored, and the results from sample surveys can be grossed up to population estimates.

- ▸ Provides the information for assembling mailing lists, dispatching questionnaires, monitoring responses and contacting units in the event of non-response.

10.118. Such functionality enables SBR and survey staff to monitor their operations and to identify areas of weakness and potential quality improvements.

10.10 Examples of quality indicators by quality dimension

10.119. The particular choice of quality indicators for an SBR depends on the situation of a national statistical office in terms of survey programme, statistical infrastructure and the available resources. Examples of possible quality indicators by quality dimension are shown in figures 10.1–10.3. A much more comprehensive set of indicators developed by the National Institute of Statistics of Italy (Istat) for its SBR is detailed in annex D1.

Figure 10.1
Input phases

Dimension	Indicators
Timeliness	Time lag for each administrative source (the time lag between registration or updating of data with the administrative source and the delivery of this administrative data to the SBR)
Accuracy	Coverage of each source (measuring undercoverage and overcoverage) Completeness (of characteristics) for each source

Figure 10.2
Processing phases

Dimension	Indicators
Punctuality	Time lag (whether survey frame available according to agreed schedule)
Accuracy	Coverage and completeness (measuring the extent to which administrative data are integrated into the SBR) Extent of missed matches and mismatches in linking administrative units and statistical units
Accessibility	How readily available relevant data are to SBR staff
Coherence	Internal coherence of units and characteristics within the SBR Coherence with other registers

Figure 10.3
Output phases

Dimension	Indicators
Timeliness	Time lag for each survey (time lag in delivery of the survey frame relative to the survey reference period) Currency of the values of the characteristics
Accuracy	Coverage (completeness of the survey frame for given reference period) Completeness (of values of characteristics) Error rates (e.g. percentage of units with wrong address, percentage of units with wrong activity code) Non-response rates (as indicator of errors in contact information)
Accessibility	How readily available relevant data are to the users
Coherence	Coherence with other registers

Chapter 11
Key considerations in establishing a statistical business register

11.1   Introduction

11.1.   This chapter provides guidance on the planning, organizational, legal and technical (IT system) factors that position an SBR for success. It is intended to offer practical suggestions without being overly prescriptive. The environmental circumstances and factors within which countries build their SBRs may be vastly different. The legal frameworks for acquiring data and the access to human, financial and technical resources ultimately greatly influence how a national statistical office can proceed.

11.2.   Some key themes recur, including the following.

> ‣ Effective partnerships must be built with data suppliers, funding providers and SBR users, first, by ensuring recognition of the critical role of the SBR in delivering a coherent and reliable national economic statistics programme, and, second, by setting up robust governance structures and partner engagement mechanisms.

> ‣ Implementing and operating the SBR must be managed in a manner that enables it to focus on and achieve its primary purpose, which is to identify the population of economically active units in a country so that it can be surveyed to acquire useful economic data. An SBR can fulfil other secondary yet highly desirable roles, such as acting as a data-collection management and tracking tool, as described in chapters 2 and 8. The SBR design and implementation plan should allow for adding the corresponding components, but only once the SBR has fully matured as a source of quality statistical frames.

> ‣ The approach should be to maintain simplicity to the extent possible. Conceptual and technical complexities should be added only when they serve a practical purpose; they should never divert the SBR from meeting its larger goals.

11.3.   The purpose of chapter 11 is to provide guidance to SBR developers and managers. Section 11.2 reviews the key considerations in establishing an SBR and recommends a modular approach to the development of the SBR. Section 11.3 is about governance and organizational structure, legislative framework, funding, human resources and relations with other registers. Section 11.4 offers recommendations on the IT infrastructure and programming requirements and, lastly, section 11.5 considers the issue of data retention or safekeeping of the SBR historical information in order to satisfy operational and analytical purposes.

11.4.   The primary reference for the chapter is the African Development Bank's Guidelines for Building Statistical Business Registers in Africa (2014): chapters 13, 14, 16 and 17.

11.2   Planning considerations

11.2.1   Initial versus longer-term scoping: the importance of a modular approach

11.5.   The primary function of an SBR is as a central frame for economic surveys. This enables conceptual coherence and creates the foundation for an integrated economic statistics programme. This fundamental purpose should be the focus at the outset of SBR development.

11.6.   The longer-term vision, however, should, from the beginning, also allow for the addition of other features and components that will further enhance the value-added of the SBR. The secondary outputs to be potentially developed after the SBR has become operational as a survey frame are described in chapters 2 and 8 and comprise the following:

- ➤ A survey feedback mechanism that facilitates the update of frame information based on the information pertaining to frame-based characteristics, such as economic activity classification.
- ➤ A receptacle for tracking survey collection outcomes, response rates and other metrics.
- ➤ A module to track respondents and response burden.
- ➤ A source of register-based statistics: it may be advisable for statistics derived from SBR data to ensure the coordination of the base population, as noted in section 2.10.
- ➤ A source for national and international data exchange: this requires the usage of common concepts and methods to ensure the coherence in the units and statistics produced by the national statistical offices involved. This is important, as SBR data are used in studies comparing the structure of different national economies and other international studies.

11.7.   In particular, the IT professionals who design the SBR data structures and the overall system benefit from having a longer-term vision clearly defined in as much detail as possible. This simplifies the addition of modules as the SBR evolves. Defining all the different modules right at the beginning enables their development to be prioritized and their implementation to be phased.

11.8.   An important development task is to determine the characteristics that need to be included in the survey frames, and hence the characteristics that have to be reliably populated and maintained with data from either administrative or statistical sources.

Box 11.1
Main challenges in IT development for the SBR in Indonesia

BPS faces a number of challenges in the IT development of the SBR in Indonesia. At the planning stage, there is often doubt and scepticism. It is therefore important to:

- ➤ Engage various stakeholders actively in the process of development
- ➤ Explain the benefits of the SBR
- ➤ Cite successful examples in other countries
- ➤ Decide which surveys will be prioritized for using the SBR

During the development process of the SBR, the IT department must be fully involved to understand the concept and business process development of the SBR in the overall

Box 11.1 (*continued*)

organization. They need to follow the meetings, read the reports and monitor the progress. The system should be developed gradually through prototyping, since the concept is still developing. BPS can make it scalable by design, because at the beginning of drafting, working methods are not yet mature, and a trial-and-error process is still under way. There will be overlapping in data coverage among subject matter areas that needs to be reconciled.

In the integration with the overall national statistical office transformation process, the relationship between work units, and inter-system relationships should be considered. In the case of Indonesia, with the STATCAP-CERDAS transformation programme, the design of the SBR must be coherent and align with other business processes or systems, such as the relationship between SBR and other Corporate Statistical Infrastructures (Collection Systems, Survey Management System, Household Frame, Administrative Data Acquisition Units System, etc.) and the relationship between the SBR system and the overall integrated statistical system.

11.2.2   Key considerations for establishing a survey frame

11.9.   To reiterate, the SBR must first and foremost provide a reliable listing of enterprises (in the form of a frozen frame) on the basis of which surveys can accurately measure the economic trends of a country. Creating the SBR is challenging, as is keeping it up to date once it is in use. The challenges of creation and maintenance are greatly eased by adhering to the principles outlined below.

Do not over-extend resources in the early stages by trying to cover all types of enterprise

11.10.   While a highly developed SBR may cover a vast portion of the economic population, a new SBR must focus on covering the part of the population that is most important economically and that can be most reliably captured and reflected. The need to manage with limited human and technological resources and to use initial funding efficiently should limit the scope of the initial SBR population.

11.11.   Reflecting the informal economy, which is highly diversified and for which no administrative data exist, cannot be a focus of the SBR development project. In certain countries, however – in particular developing and emerging economies – the informal economy is important, and collecting economic data about it is a priority. Thus, as was discussed in chapter 3, although the informal sector is not the main focus in setting up the SBR, in such countries it is vital to consider to what extent, if at all, the SBR will cover the informal sector and how the economic production of those enterprises not covered by the SBR can be measured or estimated.

11.12.   It is essential to maximize the coverage in every sector of the economy to obtain reliable macroeconomic indicators. Since certain sectors are composed of numerous small enterprises, maximizing coverage in these sectors implies the addition of a large volume of records to the SBR that result in a very small contribution to overall macroeconomic indicators (such as GDP) but significant contributions in certain sectors.

11.13.   Including these small businesses is also useful as the data that they provide support policy analysis pertaining to business formation strategies, small business financing and other microeconomic issues, and greatly assist in the economic analysis of entrepreneurship and the SME sector. Thus, a progressive approach is rec-

ommended whereby a country first focuses on developing an SBR system designed to maintain large units that are economically most relevant to its GDP, and subsequently aims to develop automated processes to include the smaller units. Completing the development of the SBR in terms of coverage is one of today's common concerns in developed countries. The initial development plan for the SBR system should incorporate the flexibility to accommodate a substantial increase in coverage due to the later inclusion of a (potentially very large) number of small units.

11.14.   In determining appropriate coverage, the environment within which the national statistical office operates should be considered, along with the data that can be reliably acquired from administrative sources and the resources that are required to efficiently obtain, process, and load the administrative records.

Plan for a system that provides both live and frozen versions of the register

11.15.   As initially discussed in chapter 2, the SBR environment should include both the live register and the frozen frames:

> ▸   The live register receives updates that are instantly recorded and used for contacting the businesses.

> ▸   The sequence of frozen frames is produced from the live register monthly, quarterly or annually and is the source from which survey programmes draw their sample files.

11.16.   A frozen frame contains all the salient characteristics that are stored in the SBR for a particular unit, including its unique identifier and its size, geographical location, sector and economic activity characteristics. Survey programmes can use a frozen frame to identify and stratify their subpopulation of interest and draw samples.

11.17.   The set of frozen frames also provides a basis for period-to-period comparisons of frame quality. Point-in-time estimates can be calculated and compared with one another to examine the number of units being born, dead, reclassified and so forth. This greatly facilitates the identification of anomalies and problem records. Frozen frames are also used in research and analysis projects including, for example, business demographics and longitudinal studies such as entrepreneurship analysis.

11.3   Governance and organizational considerations

11.18.   Organizational aspects of an SBR, such as relationships with administrative data providers (e.g. the tax authority), stakeholders (e.g. the central bank) and SBR users (especially survey statisticians) are discussed in this section.

11.19.   In almost all national statistical offices, the SBR resides in an economic division or department or a statistical infrastructure division or department within the national statistical office.

11.20.   The governance and organizational structure of the SBR within national statistical offices are important – both in developing the SBR and, even more, in maintaining it and providing support for users. The SBR should be, if possible, a separate organizational unit with a dedicated manager within the national statistical office. The unit should assume the following responsibilities:

> ▸   Define and document all concepts, in line with international, national and local statistical standards.

> ▸   Plan and direct the development of SBR system processes and functionalities, including built-in business rules and system edits to prevent significant errors and enhance system control.

> ‣ Plan and implement a quality assurance programme for the SBR with the goals of assessing its quality and ensuring its continued integrity, and defining and producing SBR quality indicators.

> ‣ Identify system improvements and recommending adjustments to the training programme or procedures where required.

> ‣ Profile businesses to delineate those that are larger and more complex, and thus to properly represent their production output.

> ‣ Ensure that businesses are classified according to the standard economic activity classification.

> ‣ Create statistical units, using data from administrative sources, profiling activities and other surveys, to create a complete and unduplicated SBR aligned with the needs of 2008 SNA and other users.

> ‣ Validate new development strategies, specifications and procedures.

> ‣ Develop and deliver courses and material to educate the full range of SBR users and stakeholders, including profilers, frame specialists, analysts, coders, survey areas, and collection areas.

> ‣ Develop a certification process so that those wishing to access the SBR must first achieve an appropriate level.

> ‣ Develop an online system for all internal users, enabling them to browse and update specific enterprise structures and specific units.

> ‣ Support those who use SBR data, which includes evaluating their needs as related to surveys or analysis.

> ‣ Provide direction and support on legal aspects related to SBR data, such as access and dissemination.

> ‣ Maintain a dedicated group responsible for producing data for users and processing all data related to SBR maintenance.

Box 11.2
Example from Statistics Canada

In Canada, the SBR is a critical component of the economic statistics programmes of the national statistical agency, Statistics Canada, and is therefore held, managed and maintained within that agency. Statistics Canada belongs within the portfolio of the Minister of Industry and also has close ties to the Department of Finance, other federal departments, and provincial, territorial and local government organizations, with the aim of ensuring that the economic statistics are relevant.

The strategic direction of Statistics Canada's programme is governed by various committees and working groups across the different levels of government. The programme's governance model provides clear direction, ensures periodic review of results and enables identification and execution of adjustments to achieve expected outcomes. The governance framework also enables transparent, effective and efficient decision-making, and supports accountability and continuous improvement of the programme.

11.3.1   Relationships with users

11.21.   To understand and gather users' needs, the SBR should have a consultation mechanism. The most important users are the survey areas and ultimately, the SNA. The SBR staff should meet survey statisticians regularly to understand the changes in the economic world and corresponding requirements for SBR data. Every request from a user should be reviewed and prioritized based on its potential impact (benefits) and the availability of resources.

Box 11.3

Portion of the Statistics Act that grants Statistics Canada access to administrative data

Statistics Act (sect. 13): A person having the custody or charge of any documents or records that are maintained in any department or in any municipal office, corporation, business or organization, from which information sought in respect of the objects of this Act can be obtained or that would aid in the completion or correction of that information, shall grant access thereto for those purposes to a person authorized by the Chief Statistician to obtain that information or aid in the completion or correction of that information.

11.22.   As the SBR is a single source serving multiple users, balancing requests within the limits of the resources available can be a challenging. The SBR manager must always be aware of the role of the SBR within the context of the larger statistical programme.

11.3.2   Legislative framework

11.23.   As stated earlier, access to administrative records, such as corporate tax returns, business registrations, payroll deductions and VAT remittances, is fundamental in building a centralized SBR. Many countries legally require the provision of such administrative data to the national statistical office for the purpose of compiling official statistics, and this authority, where it exists, must be fully leveraged to build an SBR.

11.24.   In implementing the legislation, or even in the absence of it, detailed agreements, usually formal memorandums of understanding or service-level agreements, should be signed with the administrative agencies, as discussed in section 6.2.1. Memorandums of understanding are crucial because they establish a framework of general rules and procedures for interdepartmental data exchanges, and they specify data protection measures. More specifically, a memorandum of understanding should clearly spell out the terms for acquiring the administrative data; it should stipulate the types of data required, the likely treatment of the data and the pre- established acquisition schedule; and it should also specify the data security measures and the means of transmission to be used.

11.3.3   Funding and development phases

11.25.   The funding model for the SBR is itself an important aspect of its governance and decision-making process, and should be carefully considered at the outset.

11.26.   The costs of creating and maintaining the SBR are not likely to be recovered on a fee-for-service basis, as the SBR is a public-good type of resource used by multiple programmes for diverse purposes. Some individualized activities, such as the preparation of special data requests for individual external users, may yield revenues to offset

the incremental costs of serving those users. Benchmarking of cost structures may give useful information relating to cost-efficiency. Developing and maintaining the SBR are generalized expenses, however, the majority of which are likely funded from the public purse, for example, as part of the national statistical office's budget.

11.27. The key questions are thus: who should make the budget requests? And who should control the budgeted funds once they are received?

11.28. The precise answers depend on the organizational governance structures in place, but the overall objectives should be the same, namely, to ensure that funds are allocated (and guarantee a sustained functional budget) so as to meet the larger objectives of the economic statistics programme in an optimal fashion. The manager of the SBR should be able to make specific budget decisions within the framework of this larger governance structure.

11.29. Developing and implementing an SBR consists of three distinct phases from a funding perspective: pre-build, development, and post-build operations. It is important that the last mentioned be included the development funding plans.

Conceptual development (pre-build) phase

11.30. This phase requires initial funding and resources for the following pre-build tasks:

> ‣ Defining the purposes, uses and roles of the SBR, namely, the data and services that should be produced to meet users' needs: to this end, consultations with users and stakeholders should occur from the start through all phases.

> ‣ Outlining a clear overall picture of the future SBR.

> ‣ Defining the initial scope of an SBR development project.

> ‣ Defining the initial coverage of the population of enterprises.

> ‣ Determining a method or methods to derive the statistical units.

> ‣ Analysing and determining all the data inputs needed to populate the live register and to create a coherent, accurate and timely set of frozen frames.

> ‣ Obtaining access to the input data sources; this may require the drafting of memorandums of understanding between the national statistical office and other departments; and negotiating legal, ongoing data transfer agreements with organizations that will supply data, such as the authorities responsible for tax data.

> ‣ Determining the availability of, and analysing, the necessary stratification characteristics.

> ‣ Defining activation rules, to indicate what signals or specific characteristics determine when an enterprise is to be considered active.

> ‣ Defining the most viable means of receiving business inactivation and cessation information.

> ‣ Identifying and determining the availability of administrative data from various sources; initial and ongoing cost is greatly influenced by the accessibility and usability of data input.

> ‣ Determining if there will be (or should be) an economic census as a source of input which itself depends on the availability and accessibility of administrative data.

> ‣ Outlining the processes that will assess the quality of the data and validate and treat them before they are loaded into the SBR.

> ‣ Defining components for future development, using the modular approach as discussed earlier in this chapter, such as a module to track respondents and the response burden.

Development phase

11.31.   Development phase funding depends on the decisions made during the prebuild phase. A modular development approach is advantageous because development can be phased in and compartmentalized, module by module. For a fully operational SBR, the following modules should be funded and fully functional for the initial SBR:

> ‣ Live register, covering targeted population

> ‣ Batch load and update processes

> ‣ Online tools supporting SBR maintenance

> ‣ Frozen frames, outputs for surveys

> ‣ Quality assurance

Post-build operations phase

11.32.   The SBR should be seen as a continuously evolving entity within the national statistical office. A long-term vision supported by senior management is needed. A senior steering committee should oversee the needs and usage of the SBR. This ensures that the SBR will evolve in accordance with the agency's general data requirements and functionalities/roles. A 10-year investment plan should be outlined as input to preparation of the funding schedule. In addition to the funds reserved to support the ongoing production and system maintenance of the SBR, this long-term plan should be reviewed annually and adjusted to new business requirements and the current situation of the national statistical office budget.

11.3.4   Human resources

11.33.   The human resources allocated to the SBR depend on the financial support that the national statistical office is prepared to dedicate to developing and maintaining an SBR. Investing in a high-quality SBR that is timely, accurate, coherent, and user-friendly will result in lower costs and better quality in other parts of the economic statistics programme, for example by facilitating more precise sampling methodology, lowering collection costs and respondent burden, improving response rates and providing higher quality estimates. To the fullest extent possible, the human resources allocated to the SBR should be dedicated to the SBR: in other words, should not have any other function.

11.34.   Regarding the organizational structure of the SBR, there is no international standard, or even commonly used practice, among national statistical offices. Accordingly, each national statistical office has to determine its own particular SBR structure and the place of the SBR within the economic statistics infrastructure and programme as a whole. Factors to consider in this determination include the following:

> ‣ Numbers of people working in the national statistical office and economic statistics programme – evidently the SBR cannot expect to have more than its fair share of the human resources.

> ‣ Number of surveys that the SBR is servicing, or will service.

‣ Degree of centralization or regionalization of the national statistical office and scope or need for regional business registers.

‣ Sophistication of computer systems supporting SBR operations.

Ongoing operational requirements and organization

11.35.   The operational structure of a completed SBR and its corresponding funding structure should be clearly set out in advance to prevent the build project from creating an ineffective or underused SBR. Although many alternative groupings of the functions are possible, figure 11.1 and the following paragraphs present a feasible structure involving three SBR sections or organizational work units, with distinct roles: operations and data maintenance; data management and quality assurance; and IT and systems.

11.36.   Each section may be expanded, contracted or modified as need be to take account of the particular circumstances of the national statistical office, in particular the resources that it has at its disposal, and the functions that the SBR performs. The sectional structuring has been devised to delineate clearly the essential tasks, responsibilities and funding commitments. Opportunities for efficiency and collaboration between the sections can be pursued.

Operations and data maintenance section

11.37.   The staff of this section undertake all the basic tasks associated with operating an SBR, including processing administrative data, processing frame data feedback from surveys, performing profiling activities, producing frozen frames and training staff and users. Securing and committing full-time personnel to these tasks is essential to ensure that the SBR data remain relevant and accurate. The size of the section depends on the coverage of the SBR at its outset. Its structure may be expected to change as SBR coverage and content improve.

Figure 11.1
Operational structure of an SBR
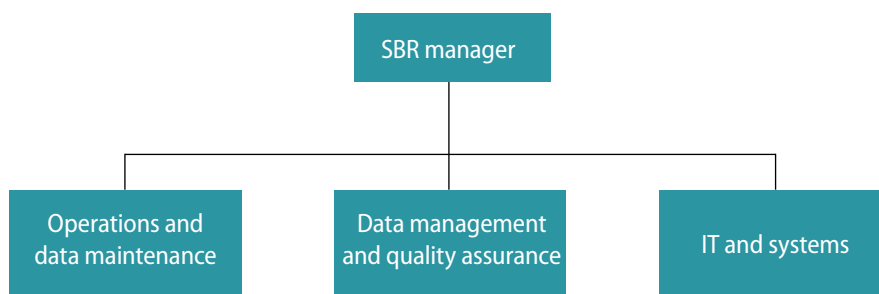
11.38.   A realistic, well-funded and evolving maintenance strategy is a crucial aspect of the SBR. For example, in the case of profiling, the exact nature of the signals and processes that prompt a profiler to review a particular enterprise structure should be reassessed over time. When an SBR is new, the section might simply schedule regular profiling of the largest enterprises. Subsequently, it might target profiling operations

more specifically to those large enterprises for which data collection is proving most difficult.

11.39.   The section should comprise:

‣   Specialists in SBR concepts and business structures, to perform profiling

‣   Specialists in administrative data sources and processes

‣   Industry classification specialists

‣   Operational staff

11.40.   The section should be responsible for:

‣   Acquiring and processing administrative data

‣   Processing frame data feedback from surveys

‣   Performing profiling and classification activities

‣   Investigating and correcting errors and inconsistencies that are identified by users or the quality assurance section

‣   Producing frozen frames

‣   Training SBR staff and users

Data management and quality assurance section

11.41.   While the SBR is being established, and after it is established, the scope of the staff that are initially responsible for analysing the incoming data should be extended to assessing quality requirements and developing a quality assurance strategy. Consistency checks should be built into the system to improve system control and prevent errors. A person, or a team, ideally with a good understanding of SBR concepts and of enterprise structures and strong technical skills in validating data and data processes, should analyse frame data and investigate potential quality issues.

11.42.   This data management and quality assurance section should be maintained as part of the ongoing operation of the SBR. The team should comprise:

‣   Data source specialists, to create specifications and use cases for the development of business rules and automated processes for acquiring and processing administrative and statistical data.

‣   Specialists in SBR concepts and business structures (having experience in economic statistics and survey cycle is relevant).

‣   Survey interface specialists.

‣   Data analysts.

11.43.   This team should be responsible for:

‣   Providing specifications of the characteristics to be populated in the database tables, maintaining the content of the tables and submitting new or modified specifications as needs arise and shift.

‣   Performing acceptance testing of changes made to the SBR systems, interfaces and processes.

‣   Assessing data from administrative sources.

‣   Conducting internal analysis and coherence checks on frame information.

‣   Reviewing requests from survey statisticians and other SBR users (including SBR staff themselves) for improvements or modifications to SBR metadata, characteristics and outputs, prioritizing these requests and creating specifications for future development and improvements.

> ▸ Liaising with survey staff to ensure that they understand SBR processes and outputs, and that they are submitting and receiving files from the SBR as needed.

> ▸ Ensuring the quality of frozen frames before they are released to users.

> ▸ Identifying significant errors.

> ▸ Ensuring that updates are timely, coherent and accurate.

> ▸ Documenting concepts, processes or coverage changes over time.

IT and systems section

11.44.   Maintaining the SBR requires dedicated IT staff. The IT and systems section ensures that the systems and software needed for extracting data are properly programmed and optimized. The section also maintains, and possibly advances or further develops, the graphical software that enables SBR staff easily to access and update the SBR content. This may initially be restricted to packaged server software, such as Microsoft SQL Server Management Studio. In addition, the section ensures that certain derivation tasks are programmed and run correctly.

11.45.   At least one database administrator is required. The administrator interacts with the database management system processes and tables that make up the SBR and ensures that the tables are accessible and available for production of the outputs. The database administrator is responsible for ensuring that the database is functional and also ensures that the database can be updated and queried as required.

11.46.   The number of dedicated IT personnel required for the implementation and maintenance of the SBR depends on the volume and nature of the functional and non-functional requirements for the implementation of a specific SBR. They must also be competent to handle specific IT requirements, to develop and maintain the SBR application and to test newly developed features. The number required could range from one or two dedicated persons for a simple SBR to a team of experts during the course of a complex implementation.

11.47.   The types of IT specialists required include:

> ▸ Architect

> ▸ Database administrator

> ▸ Developer, with knowledge (as needed) of software development tools and languages such as .NET, Java, SAS, T-SQL, PL/SQL

> ▸ Information analyst

> ▸ IT project manager

> ▸ Systems analyst

> ▸ Tester

11.48.   In smaller teams one person may perform more than one type of task.

11.3.5   SBR training

11.49.   The SBR manager is responsible for developing and maintaining a training programme for national statistical office staff who will work with the SBR. The training programme should be designed to be easily divided into separate modules, which can focus on specific needs of different user groups.

11.50.   As a first priority, the training programme should be designed and implemented in such a manner as to ensure that all users understand the basic concepts and

information available in the SBR. This training should be delivered to all users who are planning to use or interact with the SBR.

11.51.   Second, all users who are planning to use the SBR system interface should be trained in the system's functionalities.

11.52.   Third, a more intensive training programme should be developed and provided to the staff that conduct profiling activities and perform frame updates. This is particularly important in ensuring that profiles of complex business structures are completed and organized properly so that the appropriate statistical elements from which samples will be drawn can be consistently obtained.

11.53.   Finally, if the SBR is accessible through an interface by the survey collection staff, training should be developed and given to interviewers who directly contact businesses in the course of survey collection activities. This part of the training programme would focus on the maintenance of the respondent (business contact) information and should ensure that interviewers are trained to gather the necessary information to update the frame information directly or provide the relevant information to the SBR maintenance team or profilers.

11.54.   It is likely that other specialized training may need to be developed, depending on the variety of SBR functions available, the types of access that can be granted and various user roles allowed. There is also the possibility of aligning training with a user certification process which gives different permissions to access the various areas and functions of the SBR. Ideally, certification would require both attendance at the necessary training sessions and successful completion of written tests to ensure that the content is adequately grasped.

11.55.   In summary, the SBR training programme should comprise specific courses and training material that address the needs and requirements of all users, ensuring that they have all the information that they need to use the SBR system with the appropriate level of access and permissions.

11.56.   It is important to keep training materials and courses updated to reflect developments in the SBR and to meet the evolving needs of users. New users should be able to receive support in the form of materials and training that they need within a reasonable time frame. Training materials and other documentation supporting the training should include:

> ‣ SBR concept definitions, including types of units, attributes and relationships
> ‣ Interface functionalities
> ‣ Industrial classification
> ‣ Update guide and procedures
> ‣ Database metadata and frozen frame record layout
> ‣ Glossary

11.4   IT considerations

11.57.   This section provides notes and recommendations on the IT infrastructure and programming requirements for the build phase of an SBR system. Cost and resource requirements may be inferred.

11.58.   When establishing an SBR, there are many possible technologies. The choice should take into account scalability, cost and maintenance. The technology should be flexible enough to evolve in line with new requirements.

11.59.   As there is no international standard or even commonly used practice among national statistical offices regarding the design of an SBR system per se, general international IT standards and guidelines, such as The Open Group Architecture Framework (TOGAF) developed by The Open Group should be considered. The main consideration is to develop an SBR system that fits within the national statistical office IT architecture and that is as compatible as possible with other systems like the administrative data acquisition systems and the business survey collection systems.

11.4.1   Project management methodology

11.60.   The development and implementation of a new SBR is a significant undertaking and should be managed as a project. If the national statistical office has a defined project management methodology, then it should be applied. If there is no existing organizational standard, then widely accepted international standards such as PRINCE2 and PMI/PMBOK should be evaluated with the goal of selecting the one that is most suitable.

11.61.   As a general rule, the level of rigour in implementing the project management methodology should be scaled to the size of, complexity of, and risk associated with the project.

11.4.2   Software development methodology

11.62.   The development and implementation of the system required to support the establishment of the SBR is always a substantial part of the overall effort required. Many organizations have IT standards and, of course, these should be followed. Where this is not the case, an appropriate methodology should be adopted. Modern methodologies place a high focus on agility, demonstrating value to stakeholders quickly and regularly, managing change, proving the technical architecture early, managing risk continuously and establishing frequent checkpoints (quality gates) for the evaluation and realignment of scope, cost, schedule and quality.

11.63.   IT development can be implemented in two phases:

(a)   Acquiring an initial database infrastructure: keeping in mind the modular approach to growing the SBR, the infrastructure should be expandable and flexible;

(b)   Programming and process development: this is required in order to create:

›   The tables that make up the database.

›   The main statistical outputs of the frame (sampling frames, sample files, and so forth).

›   The applications, programs and tools required to query and update the database to support business workflows and interface with other systems, for example, the administrative data, survey and collection systems.

›   Data and metadata transfer processes in a standard format (e.g. Statistical Data and Metadata eXchange (SDMX), to facilitate data and international comparability.

11.64.   As with project management methodology, the software methodology should be scaled to the size, complexity and risk of the project.

> **Box 11.4**
> **SDMX as a tool for the exchange of data among different national statistical offices to enrich their statistical business registers**
>
> It is essential to identify useful information that can be internationally comparable. Thus, for the SBR, the following may be considered: the number of businesses existing in a given year; births and deaths of enterprises; growth by sectors of economic activity, at the macro level, and, at the micro level, the complete or partially integrated directories of the SBR, considering the legislation of each country, including data such as legal name, address and location. This is a useful tool because it enables comparability and information exchange among countries in a more efficient way, by ensuring the granularity needed for studying the global production and the role of the multinational enterprises.
>
> Usually, for the application of SDMX, the construction of the data structure definition is required with the list of fields, concepts, codes and its connection with the database, covering all the information that could be useful among countries and international organizations. The data structure definition is key to ensuring an efficient interchange of information based on standardized databases.
>
> For further details, see the SDMX website https://sdmx.org/.

11.4.3   Solution architecture

11.65.   The requirements for an IT solution are commonly split into two groups; functional requirements and non-functional requirements.

11.66.   Functional requirements indicate what the system is expected to do, namely, the set of functions and user tasks that the system must be able to implement. A function can be described in terms of inputs, behaviour and outputs. The scope of the SBR coverage, the number of characteristics (real, estimated and derived), and the required interfaces and interactions with other systems are, along with other items, reflected in the functional requirements of the system.

11.67.   Non-functional requirements indicate overall requirements that do not pertain specifically to functions of the system. Examples include compliance with the organization's technical standards, performance, scalability, security, availability, accessibility, quality and usability.

11.68.   Functional and non-functional requirements are necessary inputs to decisions that need to be made regarding the solution architecture. The architect considers patterns, practices, tools and technologies to arrive at a design and system architecture that optimally satisfies the requirements. Most modern, complex solutions are implemented with a layered architecture, a simple generic example of which is provided in figure 11.2.

Figure 11.2
Layered architecture

Presentation layer	Implements the user interface and manages user interaction with the system
Service layer	Exposes interfaces and system functionality to other systems, and may also be the boundary between the presentation and business layers
Business layer	Implements the core functionality and business logic
Data access layer	Implements access to and interaction with data stores

11.4.4 Database

11.69. There are a number of options, including a relational database management system (RDBMS), an object-oriented database management system (OODBMS), a key-value store, hierarchical database system, flat files, or even a spreadsheet-based solution. An RDBMS is the predominant choice for a core piece of infrastructure such as an SBR. It is recommended on the basis of the following strengths:

> ▸ Extensive capacity to implement security
> ▸ Extensive capacity for scalability
> ▸ Extensive capacity for support of concurrent users
> ▸ Capacity to enforce referential integrity and improved data integrity
> ▸ Efficient storage facilitated by normalization and lack of data duplication
> ▸ Efficient application code
> ▸ Flexible and standardized query language
> ▸ Ability to be extended for future requirements

11.70. A relational database enables proper segmenting of the areas that make up the whole, which in turn enables a proper maintenance strategy. The database should be scalable, as there may be an increasingly large number of observations to process.

11.71. SBR data should be stored in a purpose-built database that can support all the types of units in the economic units model, along with their interrelationships and characteristics, and that has appropriate confidentiality and access provisions.

11.72. There should be provision in the database, or within an associated and controlled environment, for storing and accessing data generated by all SBR functions, including snapshots, frozen frames, survey control files, respondent reporting obligations and statuses, and respondent burden.

Frame, collection and respondent burden modules

11.73. In its primary capacity of providing frames for surveys, the SBR should be seen as a database that stores both the frame data (including unique identifiers and stratification characteristics) as well as data about the collection processes. The required collection data include the units sampled, information on how and from where data about these units are to be collected and the outcomes of the collection efforts. These data are input to the SBR respondent burden module that provides an agency-wide view (in other words, across economic statistical programmes) of the reporting burden that the national statistical office places on enterprises, and a record of any efforts to mitigate this burden, for example by using tax data in place of survey data.

11.74. A key benefit of using a relational database as the foundation of the SBR is that tables can be created, developed and maintained apart from one another. This provides great flexibility regarding resource requirements and allocation, the sophistication of the various tables, and the capacity to build on different modules at different rates. Thus, initial resources should be devoted to developing the core SBR tables and its main outputs, with simple collection tables to track which units in the SBR are sampled most heavily. Future resources can be allocated more efficiently as usage is tracked. The respondent burden module may be developed at a later stage.

11.4.5   Database management system options

Relational databases and SQL

11.75.   The modular approach to building the SBR is aided by using a relational database structure that enables creation of different tables and modules in isolation from one another. Using an RDBMS solution such as Microsoft SQL Server, or Oracle, offers other benefits too, including predefined software server applications, like SQL Server Management Studio, and international standards for querying the database using Structured Query Language (SQL). When appropriate and cost-efficient, components, codes or architectures may be shared or reused.

11.76.   Figure 11.3 below, represents two tables from a relational database model. The first table (the statistical unit table) contains the statistical unit's unique identifier with a few basic characteristics of the units, and the second table (the relationship table) contains the parent identifiers of each unit.

Figure 11.3
Simplified data model example of two tables

Database cost considerations

11.77.   Functional and non-functional requirements for the SBR can have significant cost implications at the database layer in terms of licensing, amount and type of storage required, quantity and capacity requirements of the database servers, and levels of effort required by the database administrator and other IT support staff. Here are some items that need to be considered.

> ▸ Functional requirements: coverage levels, number of characteristics, scope of functionality implemented (such as number of subsystems, screens and business processing logic), and interfaces with other systems; these requirements have a direct effect on the number and size of database tables required, and therefore on the storage required.

> ▸ Non-functional requirements: performance, load, availability and number of environments; these requirements have an effect on the cost of the IT hardware required to implement the solution.

11.78.   An enterprise class RDBMS enables the addition of servers and databases as dictated by resource availability and growth.

11.79.   Another benefit of a relational database is that users can link tables across several databases to perform analyses and track data.

11.80.   When possible it will be cost-effective to reuse database code, for example, for store procedures.

Simple database solutions

11.81.   Compared with enterprise class RDBMS solutions, there are other, lower-cost options for simple database solutions. The reduced costs are accompanied, however, by loss of functionality and flexibility (for usage and future growth) compared with that offered by a relational database. For example, MS Access can be used to run a limited SBR, but with several shortcomings:

> ‣ Inability to programme sophisticated data constraints
> ‣ Limitations on how users can query and update tables simultaneously
> ‣ Constrained ability to proceed with a modular approach that enables gradual introduction of databases and tables
> ‣ Limited security controls

Database construction and organization

11.82.   Designing database tables involves deciding which tables will house which characteristics, and the manner in which these tables should interact. The IT team should also determine how data are backed up and secured, and should create a set of database roles (e.g. "write access") to be assigned to users, as needed.

11.83.   Another important task in table design is creating unique identifiers for the records in the SBR, both to identify specific records and to identify relationships between the records. Creating and maintaining hierarchical relationships is essential as they enable users to denote the parent–subsidiary relationships of enterprises, and the operational links between enterprises and establishments.

Advantages of relational database: quality assurance and data coherence processes

11.84.   Another advantage of a relational database system is that data constraints can be placed on tables. This ensures that inputs, either from statistical or from administrative sources, can be required to meet certain criteria in order to be loaded. This in term enables automated quality checks on data when they are loaded. Furthermore, dividing different types of information into multiple tables allows for data segregation and thus analysis can be still further atomized.

11.85.   The IT development team should be contracted to insert appropriate data constraints. As well, they should design data coherence programs that can be used to determine outliers and validate the internal coherence of information, for example, that the revenue of an establishment is not greater than the revenue of the corresponding enterprise. Constraints on data should also be implemented as a control on the quality of the data received and loaded from external sources.

Box 11.5
Example from Canada on identifiers

In Canada every business that registers with the Canada Revenue Agency is assigned a single business registration number (SBRN), commonly known as the BN, or Business Number. Thus, it is essential, first, for SBR staff to have a clear understanding on how to relate and link the BN to statistical units and, second, to include the BN within the SBR as a key identifier.

11.4.6   Programming requirements

Graphical user interface and database server software

11.86.   The manner in which users interact with the SBR is a crucial decision in the build process. In deciding whether or not to develop a tailor-made graphical user interface , two groups of factors should be considered: first, the types of users and their requirements; and, second, the relationship between the SBR itself and its outputs.

11.87.   Users of the SBR include:

> ‣   SBR staff whose task is to ensure that SBR information is accurate and up to date.

> ‣   Survey statisticians and analysts, who want to review the enterprises and corresponding structures in the SBR that affect their own statistical programmes with a view to updating their particular populations.

> ‣   Sampling statisticians and other statistical staff, who analyse their frames to select samples.

11.88.   Each group of users has different requirements. Sampling statisticians are most interested in analysing the outputs from the frame at aggregate level to help in optimizing sample design, and thus, for them, a graphical user interface is not particularly important.

11.89.   On the other hand, profilers and survey statisticians greatly benefit from a graphical user interface that enables them to easily search, browse and update individual units in the SBR. A graphical user interface also minimizes the need for all staff to understand SQL and database querying.

11.90.   A graphical user interface can, however, involve large development and maintenance costs, including those for design, coding and additional server infrastructure. Without a graphical user interface, there is still plenty of capacity to group and analyse enterprises based on the database tables. An economical approach may be initially to run database server software such as SQL Server Management Studio or SAS to access and query SBR data. In this way, the SBR can be made operational and can start producing the needed outputs more quickly and cheaply than if a graphical user interface has to be developed.

11.91.   In summary, given that a graphical user interface is not an essential SBR tool, in the initial stages of SBR development the cost benefit is tilted against introducing a graphical user interface. A graphical user interface may, however, be integrated into some standard database software. In addition, keeping in mind that survey frames are ultimately the main output of the SBR, a graphical user interface is a tool that may be worth introducing at a later stage to enable statistical staff more easily to browse, profile and update the units in their frames.

Production of survey frames and sample files

11.92.   The extraction process by means of which survey frames are produced requires relatively little programming, as it is a derivative product of the database tables themselves. The exact design and types of SBR database tables and the unique identifiers to be used should have already been determined before the design of the extraction processes. The main effort required from survey and sampling statisticians consists in determining which characteristics should be available in the frames and which in the sample files.

11.93.   Creating a data repository to archive frames and sample files is recommended, as these are the primary outputs of an SBR. Maintaining these files is important

from a data management perspective. They also offer a key input for future analysis and development work.

11.4.7   IT environments

11.94.   The IT solution should properly manage the deployment of new features. A full-scale solution may involve five distinct environments.

> *Production environment*: this should be a dedicated version of the system, including the active data that are being updated.

> *Practice environment*: this has the same code as production, but the data are not being continuously updated. This environment is typically used to perform updates that mimic what would happen in production. It enables users to check and see how their changes would affect the overall process.

> *User acceptance environment*: this enables testing of new programming functionality before moving the code into the production and practice environments. It typically has fictitious data that are developed solely to test various scenarios.

> *Development environment*: this is specific to systems programmers, who use it to test their own programming. Once code is system-tested, it can be moved into the user acceptance environment to be tested by users.

> *Analysis environment*: this is specific to analysts performing quality evaluation and simulation testing. This environment offers analysts access to a mirror image of the data available in the production environment without disturbing the production processes.

11.95.   The first environment is essential, the third and fifth environments are highly desirable. At a minimum, there should be a production environment and one other environment in which all forms of testing and analysis can take place.

11.96.   It is important to consider how all other systems link to the SBR when establishing these environments. For example, if end-to-end testing across systems is required, the SBR user acceptance environment needs to be linked to the user acceptance versions of the linked systems.

11.4.8   Programming language

11.97.   The choice of programming language is crucial and the options should be researched thoroughly. A poor choice could lead to premature redesign and costly maintenance processes. The chosen language should be sufficiently common that knowledgeable staff and training are readily available, and it should have a demonstrated long shelf life.

11.4.9   Relationships with other systems and registers

11.98.   The SBR should be established in the knowledge that it may be linked to other registers or lists. Thus, from the start, standard definitions and concepts should be used, and the most commonly used unique identifiers for each type of businesses should be included. If a commonly known identification numbering system for legal units does not exist, the SBR should create one itself.

11.99.   Developing a flexible SBR system that makes it possible to supplement and maintain the various concordances between identifiers is important. It facilitates link-

ing the SBR to satellite systems that create and maintain their own identifiers. A module to store and maintain specific variables of partners or SBR users will help to further facilitate interaction with other systems.

Box 11.6
Example of database structure from Statistics Canada

In Canada, the SBR is now maintained in about 30 databases on five separate servers. The online transaction processing production database is a separate database and server that can only be accessed and updated using the Business Register System application. A snapshot of the frame is stored in a separate database on a separate server, so that analysis, quality assurance and reporting activities do not affect the performance of the online transaction processing. Likewise, the Survey Interface Repository, which enables survey areas and systems to access information required for sampling and analysis of survey collection information, is stored on a separate server and database so that its workload does not affect the performance and scalability of the online transaction processing.

11.4.10   Establishing a unique identifier for statistical units

11.100. The establishment of unique identifiers is essential for accurate maintenance of the SBR. At the technical level, it is essential to have a unique identifier to load and maintain each unit within the SBR database. The sequential assignment of unique identifiers is effective. It is even more effective when the unique numbers are created and managed centrally and are used by all statistical systems and processes throughout the entire statistical organization. Having SBR-assigned identifiers is also a good means of reducing the risk of inadvertently disclosing confidential microdata by use of easily identifiable and recognizable information such as the tax number or the business name.

11.101. Given the crucial role of unique identifiers for the whole system, great care should be taken when creating the identifier system. Changes later on will be difficult to make and may have significant impacts on the statistical systems and processes.

11.102. The following are some of the key elements to consider when creating a unique identifier for the SBR:

> Create an identification numbering system for each statistical unit, no matter what type. Multiple business numbers lead to added complexity in SBR systems, and introduce risks such as duplication and omission of statistical units.

> Use a non-confidential identifier to facilitate the statistical processing. This should not contain any information about the business. All the statistical processes, such as edit, imputation and estimation, can use this non-confidential identifier in their processes.

> Ensure that these unique identifiers have no inherent meaning (other than, possibly, indicating the type of unit) and are only be used for statistical purposes. The name or similar descriptions of a unit should not be used as an identifier. Nor should the identifier contain any information on the relation between units, such as that establishment X belongs to enterprise Y.

▸ Ensure that the unique identifiers cannot be reused. Ensure also that the length of the identifier is sufficient to account for the total number of statistical units that the SBR may ever contain over time, including dead units and units of all types.

11.103. The following technical considerations should be kept in mind in the generation of a unique identifier:

▸ In order to minimize errors, consider including a check digit function, especially if the unique identifier will need to be manually captured (e.g. through the mode referred to as "heads-down data entry") by SBR users when they enter information about specific units. An example of how to calculate a check digit according to Modulo 11 may be found in annex E2.

▸ If the identifier is always generated centrally by the SBR, use can be made of a key generator function that guarantees to generate a unique sequential number. If need be, a check digit can be added to the number.

▸ In creating an identifier structure, careful consideration should be given to the possible numbers of unique IDs that can be derived. There must be sufficient space for the current plus all future numbers of units, taking also into account intended increases in the coverage of the SBR. A necessary change in the number of digits of the identifier would entail very basic adaptations of the IT database system.

▸ Alpha characters combined with numbers may be used in order to avoid confusion with any other numeric data value on the database. Alphanumeric identifiers enhance readability of the identifier. As there are more than 10 letters in the alphabet, using an alphanumeric system significantly increases the possible number of unique identifiers. Alphanumeric data need more storage space, however.

Box 11.7
Example from Canada of creation of a unique identifier

The unique identifier of the Canadian SBR is called the "statistical number". The statistical number is composed of a letter followed by an eight-digit number. The letter "S" is used, in order easily to identify the statistical number. The combination of a letter and numbers avoids confusion with other numeric data values in the database.

In the database, there is a function called the "key generator" that generates a unique sequential eight-digit number, so no digit check is required. This function also ensures the uniqueness of the generated number. *S12345678* is an example of a statistical number generated in the SBR for the unique identification of a newly created business entity.

The key generator function is used when a new statistical unit is added on the SBR. A new unit may be added online by an employee or from a batch process. For this purpose, the key generator is invoked and it uses the statistical number table that contains all potential eight-digit numbers from 00000001 to 99999999, to determine the next available number. When the function locates the next available number, it automatically places an "S" in front of the number and assigns it to the new statistical unit. As it performs this assignment, the key generator automatically removes this number from the statistical number table, leaving only numbers that have not been assigned in the table. This technique permits multiple processes to employ the key generator function simultaneously and to have access to a new eight-digit number.

11.4.11   Tools and software for record linkage

11.104. The results quoted for the performance of automated data-matching tools and software tend to be overly optimistic. Experience has shown that, when bringing together data from two sources without a known and unique correspondence between the sets of identifiers, deterministic matching or probabilistic record linkage often yields mismatches when the matching rules are too loose and a high percentage of missed matches when the rules are too rigid. Nevertheless, circumstances may dictate the need for record linkage. No specific tools or software are recommended in this chapter. Chapter 6.8 provides some guidance in this regard. It should be mentioned that data sets may have to be cleaned before they are matched against other sources.

11.4.12   Job-scheduling software

11.105. The administrative data used to update the SBR usually require a great deal of time to process. Thus, to reduce the impact on other forms of processing, administrative data are usually processed when no one is accessing the database – typically at night or over the weekend. Job-scheduling software enables efficient management of the process of running computer jobs automatically. It monitors processing and should remotely notify technicians of any problems to be resolved before users next access the database.

11.4.13   Documentation

11.106. Documentation related to the system is needed to ensure the long-term functioning of the SBR. It helps staff to identify and understand the changes that they need to make from time to time.

11.107. The IT team should invest significant time in fully documenting each module and process, both during initial development and as the SBR evolves. The documentation should be sufficiently detailed for a new programmer to continue where a previous one left off. Wikis enable free-form information entry over time.

11.5   Data retention

11.108. The SBR data-retention strategy should be articulated in accordance with operational and analytical needs and should begin with the determination of how changes made to the SBR will be tracked and what historical information will need to be kept.

11.5.1   Tracking changes

11.109. The following paragraphs summarize briefly the options to be considered for the safekeeping of historical information, such as the use of logs, the creation of snapshots and frozen frames on a regular basis, and the recording of the effective date (in other words, the time that an event really occurred), and also the update date (the date on which the event was recorded on the SBR).

11.110. In determining the strategy, a key factor is how changes to SBR data are handled. One possible approach is always to add data and never to replace data. With this approach, the (new) value of a characteristic is recorded together with the date and time when it was recorded. The old value (if any) of the characteristic and the date and time when it was recorded are retained. This approach enables the creation of a view of the SBR as at any past date and time. In a more sophisticated approach, the date and

time from which a new value is effective is recorded in addition to the date and time of the update.

11.111. Another, much simpler approach is to take periodic snapshots of the database and to keep these for as long as seems necessary. To satisfy all operational and analytical needs, using both approaches is suggested.

11.5.2   Frequency and content of snapshots

11.112. Insofar as snapshots are used, data retention should be frequent enough to offer users a reasonably comprehensive view of the past. Monthly snapshots of all statistical units and their characteristics going back two years would likely offer users sufficiently frequent data for the majority of their immediate needs. Where data retention for analytical purposes is concerned, a set of frozen frames may suffice for most users. Some national statistical offices retain these files indefinitely, as they are a valuable source of information for longitudinal studies.

Box 11.8
Example from Statistics Canada on register snapshots

At Statistics Canada a complete copy — called a snapshot — of the SBR database (live register) is taken just prior to the first day of every month. A generalized survey universe file, in other words, a frozen frame, containing every statistical unit, is created from the snapshot every month. Although frozen frames are primarily used for sampling, normally soon after their creation, they are retained for an extended period for analysis purposes. The table below shows the current retention period for each frozen frame.

Monthly frozen frames	Retention period
January	Indefinite
February–December	24 months

11.5.3   Administrative updates

11.113. Keeping track of updates from administrative sources is also recommended. An automated log should include the update date and source. This information may prove useful in troubleshooting issues arising with specific maintenance or update processes and also in documenting and explaining changes to SBR users.

11.5.4   Documentation

11.114. As the SBR constantly evolves in terms of its concepts, coverage and methods, the data-retention strategy must also contain documentation explaining these changes. Each file stored should have a full account of major changes attached to it, which will enable future analysts to understand the data. The quality assurance team may also use the log information to assess the quality of certain maintenance processes or sources of SBR updates. Once it is determined how changes made to the SBR are tracked and what historical information the SBR stores, then an information management directive should formalize what should be kept in the long term.

11.5.5 Information management directive

11.115. Assuming that there is an information management directive, the SBR should follow it in determining its data-retention strategy. If there is no such directive, a data-retention strategy should be articulated in accordance with operational and analytical needs, taking into account future cost and space requirements and general national statistical office policies and legislative restrictions.

11.116. To determine appropriate retention periods for specific sets of information, it is suggested that the following be developed:

- Groups for the various types of statistical microdata files
- Information management categories, such as categories of retention and documentation requirements

11.117. For example, four primary groups (each group relating to one or more phases of the Generic Statistical Business Process Model) could be created.

- Statistical microdata – design and collection
- Statistical microdata – processing
- Statistical microdata – analysis and dissemination
- Statistical microdata – other

11.118. Five information management categories could be created, as follows:

- Indeterminate retention and high-level documentation
- Maximum 20-year retention and medium-level documentation
- Maximum 10-year retention and medium-level documentation
- Maximum 5-year retention and minimum-level documentation
- Other (for a few file types that require special attention)

11.5.6 Archiving and deleting

11.119. As the SBR content continuously grows, the data-retention strategy must also contain documentation and rules on data and information that does not need to be retained (in other words, data and information that can be deleted) and data and information that do not need the same ease of access (i.e. data and information that can and should be archived).

11.120. The following is a list of business rules that should be considered when developing any archiving and deleting processes:

- *Automation*: The process needs to be automated with minimal manual intervention.
- *Scheduling*: The process needs to be able to schedule using a standard scheduler.
- *Logging:* A log of the tables (files) deleted (or flagged for deletion) needs to be recorded, centralized and accessible to all users having access to the process.
- *Reporting capability*: The process should create a report.
- *Possibility of review*: The process must allow the user to review the actions to be performed before being applied.
- *Possibility of executing report only*: The process should allow the user to execute the report portion of the process only (in other words, not apply the updates).

▸ *Possibility of resuming execution:* The process should allow the user to resume execution of the update portion of the process when the reporting portion has been executed beforehand. The amount of time spent between the execution of the report portion and the update portion should be flexible.

▸ *Possibility of retrieval:* The process must allow tables that have been deleted to be retrieved if deleted by mistake (at least the last batch deleted).

▸ *Validation of existing cycle:* The process must ensure that a previous cycle exists for all tables (files) that have been identified for deletion.

▸ *Flexible parameter modification:* The process should use the already existing files, tables and information so that modifications to concepts are automatically applied to the job. Parameters should not be hard-coded in the process.

Box 11.9

Statistical Business Registers for Improved Information on Small, Medium-sized, and Large Enterprise: project of the Asian Development Bank

Introduction

The Asian Development Bank aims to establish and improve SBRs in Asia and the Pacific through its capacity-development and knowledge-sharing technical assistance projects. These efforts involve two main aspects: establishing each participating country's SBR database system, and providing relevant training to the SBR staff of the statistical office.

The Bank developed an SBR system through its in-house developers, designed to be deployed and operated by the different participating national statistical offices in order to jump-start their SBR processes. Under its SBR project (R-CDTA 8594: Statistical Business Registers for Improved Information on Small, Medium-sized, and Large Enterprises, 2014–2018), the Bank has deployed the SBR system to the national statistical offices of Bhutan, Cambodia and the Lao People's Democratic Republic, and facilitated events for SBRs in the region that also included the national statistical offices of Malaysia and Sri Lanka, along with international SBR experts.

The development of this system entailed detailing common features while harmonizing the unique requirements of each country. When adding system features for use by one country, the other countries must have the option of disabling and hiding them, so that they do obstruct their use of the system. Provisions to customize the system will allow the system's code to be reused across the different countries. National statistical offices may configure their SBR system to include locally relevant variables and remove unnecessary variables and features. The system was also designed to allow national statistical offices to provide translations of the system's interface, allowing end users to operate the system in the language with which they are most comfortable.

Software development methodology

The project was initially developed in a traditional so-called "waterfall software development lifecycle". In this model, software development goes through a linear sequence of phases and the project should only move forward when the preceding phase has been completed and verified. The rationale is that time spent early in the software development cycle can reduce costs at later stages

Figure 11.4
Software development lifecycle
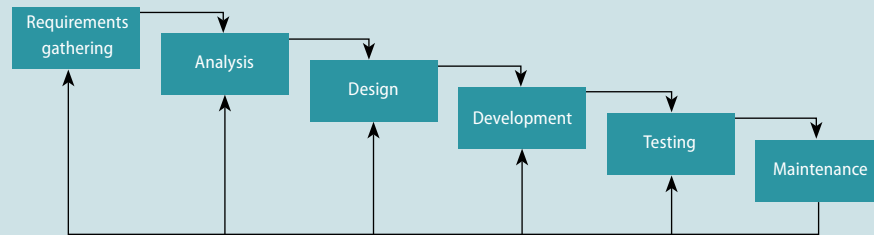
Ideally, after the system has been fully developed and is cleared for release, the developers will then shift their focus to optimization and maintenance to keep the system running smoothly.

As the system was deployed to countries over time, the team has since then shifted to adaptive and iterative development strategies at shorter intervals, in order to respond to the additional requirements received from multiple countries.

Figure 11.5
Adaptive and iterative development strategies at shorter intervals
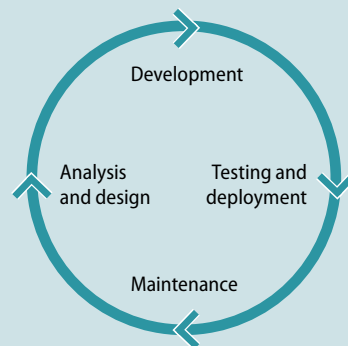
IT environments

The system was initially designed to run on a Windows-based server, using software and programming languages that form a so-called "WAMP" (Windows operating system, Apache HTTP Server, MySQL/MariaDB, PHP) stack. It was later found, however, that the SBR system's workloads perform better on similar hardware using a Linux-based set-up, in what is called a "LEMP" (Linux-based operating system, Nginx, MySQL/MariaDB, PHP) stack. Deployment, operation and maintenance of the SBR system were also significantly easier in a Linux-based system with the use of automated scripts.

Establishing a unique identifier for statistical units

The SBR system relies on a single unique identifier for the basic statistical unit, termed the "establishment identification number", or "EIN". In the process of importing or creating records, the EIN is requested from the encoder. If not provided, the system checks whether the details pertain to an establishment already existing on the system. If a matching establishment has been found, the new record is automatically linked to the EIN. Otherwise, the system creates a new EIN using an incrementing generated number, and assigns this to the new record.

Furthermore, the SBR system allows the input of several existing IDs used by source systems (such as tax identification numbers, or business registration numbers), albeit on an optional basis, as they may not be applicable to all records depending on their respective data sources. Adding these IDs is highly recommended, because they will be useful in determining duplicates or links between records through matching features that are planned additions to the SBR system. These identifiers will also facilitate future linkages of entities on external systems, as lookups can be performed to see which identifiers are linked with a certain EIN, and vice versa.

Figure 11.6
Identifiers linking records from different data sources
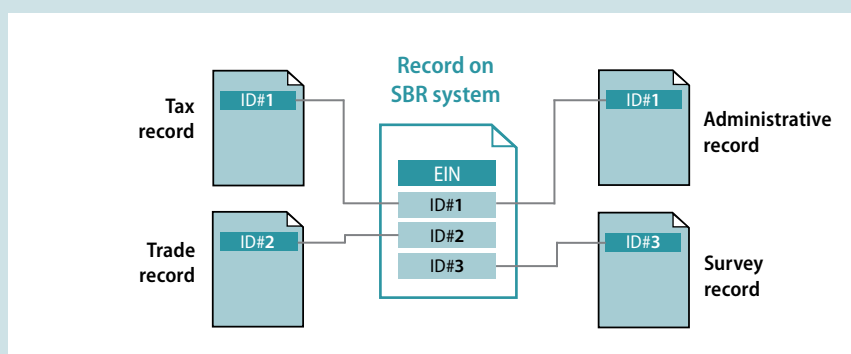
In December 2018, the Asian Development Bank published its "User Guide for ADB Statistical Business Register", which provides instruction to end users in performing typical actions from within the application's interface, and also background information about the relevant projects. The User Guide is available at www.adb.org/publications/adb-statistical-business-register-user-guide.

Box 11.10
Business register project of Statistics Norway

Introduction

For almost 25 years, Statistics Norway has worked in partnership with sister organizations around the world and, in many of these projects, assisted in the development of business registers. A common challenge for many of their partner countries has been the development and maintenance of a technical platform to host the registers. There are few off-the-shelf products in this area and many of those perform inadequately. Thus, when ECE published its "Guidelines for Statistical Business Registers" in 2015, Statistics Norway decided to make a reusable web application following those guidelines. This was done in close cooperation with the National Statistical Committee of Kyrgyzstan. The system will be licensed as open source and made available at www.statbus.org/ (the estimated first release will be in the third quarter of 2020).

Box 11.10 (*continued*)

The system is open source and free for everybody to download. It should be kept in mind, however, that this is only a technical solution and that each statistical office must pay special attention to the content of the register. The statistical office must decide on a methodology on how to use the system, arrange with external institutions to access data, assign staff for the daily work and so forth. Installing and maintaining the system requires advanced IT skills.

Technical design

The system is designed to be as flexible as possible, but some design decisions had to be made: It runs on the Microsoft Internet Information Server (IIS) and is written in ASP.NET. The underlying database can, however, be chosen freely from between the Microsoft SQL server, PostGreSQL and MySQL.

Contents of the register

Unit types

The fundamental units of the register are the legal units, but once a legal unit is entered into the system, its corresponding enterprise and local unit is created as well. In addition, there is a one-to-many relation between enterprise and legal unit (one enterprise can consist of many legal units) and between legal units and local units. The top-level unit type is the enterprise group.

KAUs and LKAUs are derived by breaking down the activities of legal units and local units respectively. The number of activities for a unit is unlimited.

Characteristics

The first version of these guidelines set out the fundamental specification requirements. Efforts were made to include everything specified while keeping in mind that the system needed to be as simple as possible. Some characteristics were added on request from the National Statistical Committee and other partners, such as whether the unit was in a (tax) free economic zone, countries of cooperation, and others.

ID elements

Ideally, each country should have a national ID number for businesses – and this ID number should be used by all relevant agencies in the public sector. Most countries do not have this, but there is often a "Tax ID", a "Statistics Office ID" and similar such codes which are used in identifying the units across different data sources. The system has the possibility of storing up to four different ID numbers:

- ‣ The actual ID, which is used by the system to identify the units.
- ‣ Statistical ID: the ID assigned by the statistical office.
- ‣ Tax ID: the ID assigned and used by the tax authorities.
- ‣ External ID: a generic ID assigned by other institutions. There is also a variable associated with this ID that identifies the institution to which it belongs.

Functionality

The aim of the project was to have sufficient functionality to cover all the needs of a statistics office, making it possible not just to modify, view and search for units but also manually to edit them:

- ‣ To upload data from a csv- or xml file: Regardless of the format of the file, xml and csv files can be uploaded by first defining a reusable template for the file, defining the order of the columns to be uploaded. Hence, if data are received on the same format on a regular basis, they can be uploaded with a few key strokes. The upload functionality also has a priority of sources: "trusted" sources, "OK" sources or "not trusted" sources, and rules whether to overwrite existing data according to the priority chosen.

Box 11.10 (*continued*)

> ▸ To create reports and statistics: As of today, the system uses third-party software (which requires a one-time licence) for the generation of reports. A set of template reports and report builders is predefined so that most reports can be generated without special knowledge. If totally new reports are needed, some SQL skills are necessary.

> ▸ To create sample frames and export the data (for further processing to draw samples).

> ▸ To conduct data quality checks: Data are flagged as "bad" according to rules set by the institution. The units can then be corrected using the modify functionality. To manage users of the system: The users (national statistical office employees) are granted write access to the register based on their area of work and/or their location. Thus, a statistician working on transport statistics can be given write access to businesses that have a corresponding activity code.

Flexibilities in the system

As this system is designed to be reused in several countries, much weight has been given to flexibility:

> ▸ Language(s): The default languages for all user interfaces are English and Russian, and switching between them is easily done by the end user. If other languages are needed, it is easy to add them.

> ▸ Activities classification: The system expects the activity classification to be in the style of ISIC or NACE (with the possibility of an additional national level added at the most detailed level), but as long as there is some kind of similar hierarchy, there is no problem with using any activity classification.

> ▸ Region classification: There can be up to five levels of regional divisions (region, district, subdistrict, etc.)

> ▸ .Other classifications (sector codes, legal forms, roles of people related to the register, size of units, etc.) must be defined prior to implementation.

> ▸ Mandatory variables: It is configurable whether variables or characteristics of the units are mandatory or not.

> ▸ Country-specific quality checks: As data quality varies greatly from country to country, some levels of bad quality can be acceptable in some countries while not in others.

Future plans and next steps

At the time of writing, the system is still being developed, with an estimated release date in autumn 2020. Implementation has already started in four countries, while there are plans for one more during 2020–2021.

There will be a user forum at www.statbus.org/ to discuss the system and post suggestions for future changes. Currently, there are no plans to develop a version 2.0 of the system, but if funding can be found, the following will be the priorities:

> ▸ Implementing requests from users posted on the SBR wiki

> ▸ Expanding the "Sample frames" functionality to draw samples, not just the sampling frames

> ▸ Creating a report builder

Box 11.11
Introducing the SBR as an integrating spine: Australia

Australia is experiencing rapid and significant changes in its data environment. The Australian Bureau of Statistics (ABS) actively participates in this environment by adding to and making sense of available data. While the traditional integrated register concept has served the survey programme well, it provides limited opportunity to leverage new data sources and support broader data analysis requirements. To better reflect the changing real-world circumstances, the ABS Business Register is being redeveloped to provide more flexibility in compiling ABS data and facilitating statistical solutions.

ABS Business Register – traditional integrated register approach

The ABS Business Register is a traditional integrated register in the form of a structured database. This works on the principle that data are maintained through an integrated data holding called the ABS Business Register, by using administrative data sources and manually collected (profiling) data. The ABS Business Register reflects and contains all the characteristics that describe the ABS Economic Units Model structure. Consistency is achieved by applying a predefined set of rules and maintenance procedures. As a result, the ABS Business Register represents the source of truth for the units used in the ABS Economic statistics program. Quarterly snapshots of the ABS Business Register are produced as outputs to inform frame populations. These frames reflect the contents of the ABS Business Register at a particular point in time.

Why the need for a new solution?

In the tradition of an integrated register, the current ABS emphasis is on the register being integrated, rather than as a means to facilitate the integration of information. The fixed structural nature of the integrated database makes it difficult to add new data items or dimensions in response to changes in the real world or the availability of new data sources.

The key purpose of the current ABS Business Register is to produce views (or frames) for the purpose of establishing collection populations to conduct surveys, rather than to facilitate the efficient reuse and integration of information for different outcomes.

Introducing the concept of the ABS Business Register as an integrating spine

To deal with the restrictions experienced with the traditional integrated register model, ABS is moving towards the approach of having its Business Register as an integrating spine. This approach is described in figure 11.7.

Figure 11.7
ABS Business Register as an integrating spine
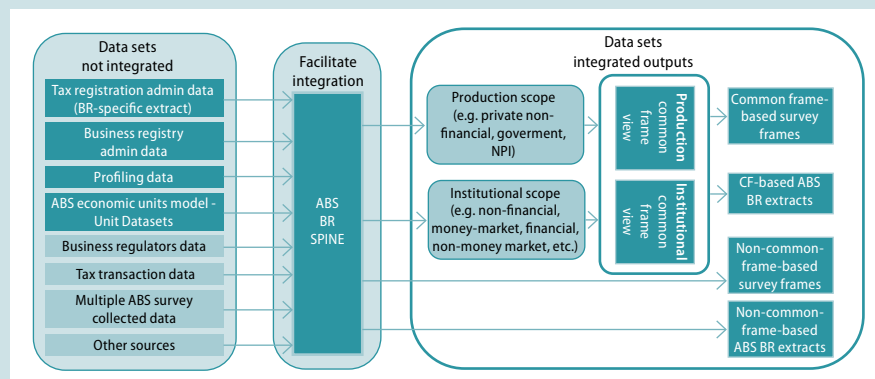
Box 11.11 (*continued*)

In contrast with the integrated register model, the integrating spine model, as set out in figure 11.7, has the potential to make use of all the data available in ABS to derive integrated outputs that could be used for multiple purposes. The components of the integrating spine model are further elaborated below.

"Not integrated" – data sets: Examples of ABS data that might be used in creating integrated outputs are shown in this block. This is not an exhaustive list as, in theory, all business-related data sets in ABS could potentially be used. These also include data sets representing the source of truth for the different administrative and statistical units in the ABS Economic Units Model. There might be a relationship between these tables, but they are not integrated on the basis of any such relationships. Each table can be used on its own and is maintained independently. Ownership of some of these tables might reside in areas other than the Business Register Unit (see darker shading for Business Register Unit ownership). Each of these tables is considered to be the source of truth, so to speak, for the purpose for which they are used: for example, ABS survey collected data, and not the ABS Business Register, are the source of truth for the survey product.

To make these tables searchable, available and clearly defined for use throughout the ABS for multiple purposes, they need to be defined and described extensively. Using the Generic Statistical Information Model (GSIM), ABS developed an ABS Information Model to describe all the data in the ABS. This is implemented through metadata that drive processes and data models.

"Facilitate integration" – ABS BR spine: In the new model the ABS Business Register spine is the source of truth for describing the scope of the ABS Business Register. It contains a complete list of the units being defined as a business for the purposes of the ABS Business Register. For the ABS Business Register, this is represented by the legal entity. Further to this, the ABS Business Register spine also contains the minimum information needed to integrate individual data sets, including the identifiers for each unit in the ABS Economic Units Model. All data in the new environment will be version controlled, so the dimensions of integration will be across data sets as well as across time (i.e. longitudinal integration opportunities). The ABS Business Register spine would be described through the same ABS Economic Units Model (in metadata).[67]

"Integrated outputs" – data sets: Unlike in the integrated business register model, here the integration is achieved by linking the input data sets and spine to produce the outputs required. This approach enables the derivation of a more diverse range of outputs and presents greater analytical opportunities.

Summary

Through implementation of the concept of the ABS Business Register as the integrating spine, ABS will be able to replicate the current existing outputs or frames, and also have the opportunity to integrate different data to replace direct collect data (e.g. with administrative data) or to use it in the development of statistical solutions.

Implementing the ABS Business Register as an integrating spine in ABS

ABS has embarked on an organization-wide transformation programme and the new ABS Business Register will be implemented through the opportunities created by this programme.

[67] Note that no integration happens on the spine itself.

Chapter 12
Topics for further work and research

12.1   Introduction

12.1.   The United Nations Guidelines on Statistical Business Registers extend the ECE Guidelines on Statistical Business Registers with a view to accommodating the needs of countries with less developed statistical systems by incorporating practical guidance and country cases applicable to a broad range of statistical systems. The United Nations Guidelines do not, however, address all the concerns and needs in developing countries regarding the identification and compilation of all businesses, and can only identify a number of emerging issues that will pose major challenges in the future.

12.2.   The continuous development of businesses, the process of economic globalization and the growing complexity of global production arrangements raise a number of questions concerning the definitions and identification of relevant statistical units of the SBR and the international comparability of statistics derived from the SBR. These developments also render data collection more complex and difficult. At the same time, the SBR has a key role to play as backbone in the modernization of statistical production and integration of economic statistics. To strengthen this role there is a need to explore how to integrate the SBR into the statistical production process as a common tool and database for all domains of business statistics, and to ensure that the SBR can be used to combine information from different sources. In addition to this, there is a continuing need to explore the use of administrative registers and other new or emerging electronic data sources and for further development of methods and concepts for the establishment and maintenance of the SBR.

12.3.   Developing new methods and good practices is resource-demanding, for which reason experiences should be shared among countries and common development projects should be encouraged. International cooperation can help reduce gaps between developed and less developed statistical systems, while capacity-building and training activities will be useful in supporting countries with less developed statistical systems. In addition, sharing methods and practices may help to improve the international comparability of SBR information.

12.4.   The topics described below build on those identified in the ECE Guidelines on Statistical Business Registers, where further work and research is recommended. Some of the topics are interrelated, since various aspects of SBR may be relevant to more than one topic.

12.2   Recommended topics by domain

12.2.1   Statistical units and profiling

12.5.   The statistical units are the basic entities in SBRs. With the globalization of economic production, the enterprise group has become an important statistical unit

that should be integrated into SBRs. Legal entities do not always meet the criteria as units suitable for statistical purposes and need to be transformed into statistical units in a consistent manner. This is especially true for large, complex and global businesses. Work in the following areas should be supported:

> ▸ Delineation of the enterprise statistical unit

> ▸ Delineation of multinational enterprise groups

> ▸ Analysis of country experiences in profiling large and complex enterprises and enterprise groups

> ▸ Sharing best practices in the coding of statistical units

12.6.   As international trade in goods and services is increasingly intrafirm trade, often organized and led by large multinational enterprises or enterprise groups, it is important to study the interlinked core production activities of the multinational enterprises and the supporting services activities to come up with a final product. These interlinked activities are commonly referred to as global value chains (GVCs). In order to support studies on GVC-specific industries, such as automotive, apparel and textile, and electronics industries, the appropriate measures of their characteristics in the SBR must be developed to better reflect their participation in the GVC, in addition to the profiling of the statistical units of global enterprise groups and the direct investment relationships between multinational enterprises.

12.2.2   SBR as the backbone of economic statistics

12.7.   It is important to develop the role of the SBR as the backbone in the production of economic statistics, up to the national accounts. The SBR fulfils this role by providing survey frames and by integrating information from different data of administrative registers) and facilitating the collection of economic data and the production of coherent statistics as part of the thrust to modernize statistical production and services. It can also facilitate the development of a standardized production process for business statistics. Countries will need to move forward and exchange experiences made in developing integrated approaches and operational designs. Activities that will facilitate this include:

> ▸ The role of the SBR in integration and analysis of multi-source data data sets.

> ▸ Integration of the SBR into data warehouse systems.

> ▸ Sharing of experiences in implementation of the GSBPM and the GSIM for the SBR database management.

> ▸ Development of efficient user-oriented maintenance strategies.

> ▸ Balancing the demands of different user groups when developing the SBR.

12.2.3   Use of administrative data sources

12.8.   The use of administrative sources continues to be a major challenge, but has the potential for gains in many countries. Administrative data in some countries are considered of low quality (in coverage and maintenance) and national statistical offices are often not given access to administrative data, owing to the absence of a legal basis or lack of cooperation by the administrative authority. In addition, there are challenges including issues with coverage and definitions, which may not be in line with

the statistical requirements, timeliness and, in some instances, confidentiality. In some countries the lack of access to administrative sources also poses an obstacle.

12.9.   Potential gains include access to sources with good coverage and detailed information that can be used directly in the SBR or in validation of SBR information, reduction of response burden and increased efficiency in the statistical office. Establishing good cooperation with the owners of administrative data sources is essential to ensure sustainable solutions and to ensure that the national statistical office is kept informed of, and exercise influence on, future changes to administrative registers. Examples of areas where work is encouraged are:

> ‣   Sharing of best practices in cooperation with administrative registers to improve the quality of the SBR and of the administrative sources.

> ‣   Sharing of best practices in cooperation with such entities as central banks and other external (governmental) organizations to ensure correct sector coding of the statistical units.

> ‣   Treatment of administrative units: linking across registers, dealing with conflicting information and working to derive statistical units on the basis of administrative units.

> ‣   Development of register-based economic censuses.

12.2.4   Use of new data sources

12.10.   New data sources and new ways of collecting data are emerging with the potential of improving the quality and coverage of the SBR, and also, in some cases, the timeliness of the register. Activities that should be pursued include:

> ‣   Sharing experiences in using new data sources, including big data.

> ‣   Sharing experiences in new electronic data collection methods.

> ‣   Legal and confidentiality issues.

12.11.   Likewise, projects and initiatives should be considered that could provide new information for the national statistical offices, such as the survey conducted by Facebook, the World Bank and OECD entitled "Future Business Survey" (www.oecd.org/sdd/business-stats/the-future-of-business-survey.htm), which analyses SMEs in the digital commerce environment. This survey is useful not only for the SBR, but also for creating and providing elements to new indicators such as those related to digital commerce.

12.2.5   Coverage of the informal sector

12.12.   Since the informal sector tends to be quite large in developing countries, its coverage in the SBR poses a serious challenge. The informal sector is defined as the set of household enterprises that have market production but are not registered in the administrative sources on which the SBR is based. Working steps to cover the informal sector could include:

> ‣   Checking all available administrative sources, including those not used for the SBR.

> ‣   Considering the conduct of a two-stage household-enterprise survey, if no administrative source is available.

12.2.6   International trade and economic globalization

12.13.   Economic globalization is one of the main challenges for SBRs. It influences the activities to be measured and data collection becomes more complex and difficult. Economic globalization is accompanied by the development of global production arrangements, often with complex ownership structures, where the challenge of the SBR is to collect and provide coherent and relevant information in a continuously changing world. At its session in March 2015, the Statistical Commission requested that a handbook on a system of extended international and global accounts be developed. The handbook was published by the United Nations in 2019.[68] This also involves new requirements for SBRs. Linkages of enterprises across borders will have to be measured and this has consequences for the classification of multinational enterprises and their foreign affiliates. Work on the following issues should be encouraged:

➤ The international statistical standards regarding, for example, the national accounts and the balance of payments, and their requirements to the SBR.

➤ The requirements for the SBR to deal with emerging global production arrangements, including, for example, factoryless goods producers.

➤ The new extended international and global accounts, and their requirements to the SBR.

➤ Pilot study for the development of a supra-national SBR to deal with multi-national enterprise groups (based on experience with the EGR). In 2015, the Statistical Commission requested the Statistics Division to develop a global group register of the largest multinationals.

➤ Linking the SBR with trade, FDI and the international sourcing of business functions.

➤ Use of the SBR for microdata linking.

➤ Analysis of enterprises in the GVC.

12.2.7   Development of new statistics

12.14.   The development of new statistics will also be a challenge for the SBR, in that information from the SBR will need to combined with information from other sources (administrative data, surveys or other statistical registers) to meet new user needs. Possibilities include combining information from the SBR with social and population data and deriving business demographic information. Questions and issues that should be addressed include:

➤ Potential of the SBR as a part of a set of interrelated statistical registers.

➤ Analysis of the effects of using different statistical units in business demography.

➤ Sharing experiences in applying geocoding in the SBR.

12.2.8   Methodological developments

12.15.   Further work is needed on various methodological issues for the establishment, maintenance and use of the SBR to take advantage of developments in techniques and data sources, and to ensure the quality and relevance of the SBR in meeting current and future user needs. Examples of particular topics where work is called for include:

➤ Improvements of economic censuses in countries where administrative sources are not available or of poor quality.

68 *Accounting for Global Value Chains: GVC Satellite Accounts and Integrated Business Statistics,* United Nations 2019. Available at https://unstats.un.org/unsd/business-stat/GVC/Accounting_for_GVC_web.pdf.

- ▸ Evaluation of database software packages that can be used for the SBR
- ▸ Coverage of the non-observed economy.
- ▸ Analysis of software services and methods for SBR maintenance that could be shared between countries.
- ▸ Analysis of various forms of frozen frames in respect of their use for business statistics.
- ▸ Analysis of methods for the creation of identifiers and check digits. Development of methods for adjustment of undercoverage.
- ▸ Development of input data validation methods.

12.2.9 International comparability

12.16. The importance of the international comparability of SBRs and derived business statistics is likely to grow in the future. Statistical offices need to consider this when developing methods and practices, while national differences in terms of legislation, institutional set-up and structural diversities place limits on obtaining international comparability. To address these challenges, work in the following areas is encouraged:

- ▸ Development of internationally comparable quality indicators.
- ▸ Development of an international quality assessment framework.
- ▸ Use and integration of international administrative identifiers in the national SBR.

Annex A
Characteristics of units by unit type

Annex A1: Enterprise group (including truncated enterprise group) characteristics

The characteristics listed in the following paragraphs apply equally to a truncated enterprise group that is the national part of an MNE and to an enterprise group that is entirely within national boundaries (all-resident group).

A1.1: Enterprise group – identification characteristics

Enterprise group – identity number

Purpose

To identify all enterprise groups (all-resident or truncated parts of international enterprise groups) and to follow their continuity.

Definition

The identity number of an enterprise group is (usually) given nationally in the SBR. It may be common to other national institutions.

As the enterprise groups are permanently buying, selling or restructuring their affiliates, the question of whether the enterprise group is the same after restructuring as before is a prominent one. General continuity rules apply for all types of enterprise groups, as discussed in section 7.3.4.

The continuity of a truncated group is based on the continuity of the global group of which it is part. The question of whether a truncated group can be continuous, even if the global group is not and even if it is split from the global group or sold to another one, remains unresolved.

Sources

The identity number of an enterprise group may be given internally in the SBR or be derived from an external (administrative) source.

Comments

As the continuity of an enterprise group may remain when the global group head changes, it is not advisable to use the identity number of the global group head as the identity number of the enterprise group.

Enterprise group name

Enterprise group – address

Enterprise group – telephone and fax numbers, email address and information to permit electronic collection of data (non-core)

Purpose

The storing of the name and address information is important for survey and profiling purposes and may have other national uses.

The enterprise group (through its head office services) is often the reporting unit for itself and even for data about other statistical units belonging to the group (e.g. enterprises).

Definition

National standards for the name and address are used for the head office, whether it be an all-resident or truncated enterprise group.

Sources

The name may be given nationally in the SBR or it may be common to other institutions (e.g. from the perspective of the European Union). Administrative sources, surveys, profiling, Internet and SBR information may be used.

Comments

The name of the truncated group should not be exactly the same as the name of the global group. It is up to the country to decide on the information that it wishes to record for each unit and address. In some countries, at the legal level the consolidated accounts may use the legal name of the group head when they refer to the whole group.

Enterprise group – identity number of the group head

Purpose

To identify the global group head (all-resident or truncated group head).

Definition

The identity number of the legal unit which is the resident group head.

Proxies

For a natural person who is not registered in the SBR as an economic operator or for a non-resident group head, a quasi-unit with an identity number can be created when necessary. In this case, it should be possible to identify natural persons separately.

Sources

Administrative sources and control links in the SBR.

Comments

If the controlling unit is a natural person who is not an economic operator, recording is subject to the availability of this information from an administrative source. In some countries recording is subject to legal limitations or banned; in other countries it poses no problem. Nevertheless, the OECD Handbook on Economic Globalization Indicators recommends the inclusion of natural persons in these cases.

The number and importance of natural persons as controlling units may depend on national legislation and thus vary considerably between countries. It would be useful to study the effect of their inclusion/exclusion on the consistency of data between countries.

This characteristic is relevant only if the global group head is a resident unit and if the unit is not a natural person. If this is not the case, a link should be made with an international register, if one exists.

The dates of commencement and cessation as resident group head must also be stored, either when the change is recorded in the SBR or preferably as separate variables when the real change takes place (if this is known).

Enterprise group – type of enterprise group

Purpose

The type is an important stratification characteristic from an economic point of view: The importance of many other characteristics depends on the type. Many economic indices may be calculated at national level by type, such as the proportion of different group types (and independent enterprises) with regard to employment, turnover per person employed, etc. The units belonging to foreign controlled truncated groups define the foreign controlled inward FATS population.

Definition

The difference between an all-resident and a multinational group depends purely on whether controlled affiliates exist in only one country or in more than one country.

The difference between domestically and foreign controlled multinational groups is determined by whether the global decision-centre managing unit is resident in the country being considered or in another country.

Sources

Administrative and commercial sources, statistical or administrative surveys, the EuroGroups register.

Comments

Types of enterprise group:

- ▸ All-resident group;
- ▸ Domestically controlled truncated group;
- ▸ Foreign controlled truncated group.

Although enterprise group type may be inferred from other characteristics it is advisable to record it separately. This is especially the case if much of the national enterprise group information is in a satellite register where the coverage may vary according to the group type.

A1.2: Enterprise group – demographic characteristics

Enterprise group – date of commencement in the country

Purpose

The date is required for monitoring the demography of enterprise groups.

Definition

The date refers either to a date when a new enterprise group is born or to another creation date of a new group (by merger, break-up, split-off or restructuring).

The date refers to the beginning of operation for a truncated group: this date can be the same as the creation date of the first affiliate in the country; but more often it refers to the date when the first affiliate is controlled (bought).

Proxies

The birth of a new group may be difficult to define in practice, if the smallest groups of no statistical importance to the country are not monitored. The date from which the group is being monitored is then used as a proxy. The approximate dates are important, however, in order to determine from which year a certain multinational group is monitored in different countries.

Sources

Administrative sources, surveys.

Comments

The creation may result from a restructuring process.

Enterprise group – date of cessation in the country

Purpose

The date is required for monitoring the demography of enterprise groups.

Definition

Cessation of a group means either death of the group in the country (dissolution of the links of control between the units belonging to the group), or (more commonly) merger with or takeover by another group, break-up, split-off or restructuring into two or more groups.

The cessation of a truncated group is registered separately from that of the global group of which it is part, because the previous demographic events can have different consequences for the global group and for its truncated part(s). Thus, the date of cessation of any kind of group is very closely dependent on the continuity rules used.

Proxies

If no exact date is known for truncated groups, the approximate date is important in order to know the situation of multinational groups in different countries.

Sources

Administrative sources, surveys.

Comments

This characteristic is recorded when the group ceases to exist or is no longer monitored. In such cases, all the characteristics become historical.

A1.3:　Enterprise group – economic stratification characteristics

Enterprise group – principal activity code at ISIC two-digit level

Enterprise group – secondary activities at ISIC two-digit level (non-core)

Purpose

These activity codes can be used for stratification, demographic and economic analysis. Secondary activities are important for observing the homogeneity of the group and, in cases where the secondary activity rather than the primary activity is the subject of a survey.

Definition

The activities are defined according to the ISIC rules. The principal activity is identified by the top-down method as the activity which contributes most to the total value added.

The same definition is applied at the global level as at the national truncated level.

Proxies

If value added data are not available, it is recommended that employment be used as the criterion. The activities performed can be based on the activity codes and number

of persons employed in the units belonging to the group at national level. Turnover is less suitable because it is more sensitive to intragroup transactions.

Sources

SBR derivation from its legal units, administrative sources, and surveys.

Comments

The main activity of any of the truncated parts may be different from the main activity of the global group.

Principal and secondary economic activities in the country can be inferred from the economic activities of the enterprises composing the group.

It is recommended that the principal activity be recorded more precisely, if possible at ISIC three-digit level.

Enterprise group – number of persons employed in the all-resident or truncated groups

Purpose

Employment can be used for stratification, demographic and economic analysis. Together with activity code, it provides information on the role of the group in the global framework.

Definition

The number of persons employed in the group is obtained by adding up from the units which belong to the enterprise group.

This procedure is valid at truncated level and at global level (provided that a global register of affiliates exists).

Proxies

The number of employees.

Sources

The SBR.

Comments

For a large multinational operating in a large number of countries, a better source for employment in foreign countries might be a specific "note" included in the annual report.

Enterprise group – consolidated turnover (non-core)

Purpose

Turnover is an important size indicator for stratification. The availability of consolidated turnover at global group, at truncated or at or subgroup level may be useful in profiling.

Definition

Turnover is defined as the revenues gained by the sales of goods and services. For consolidation rules (full consolidation) the International Financial Reporting Standards (IFRS) should be followed.

Proxies

There can be national specifications of the consolidation rules, such as the Generally Accepted Accounting Principles (GAAP) of the United States of America.

Sources

Administrative sources, surveys, profiling.

Comments

It is difficult to consolidate the turnover of different enterprises in different activities and the consolidation rules are not the same everywhere. In addition, turnover from different sources may not be consistent.

It may also be possible to aggregate the turnover of legal units belonging to the enterprise group, but this would include internal turnover which does not correspond to market prices when based on internal prices. Thus, the storing of any figure other than the consolidated turnover figure may be confusing and is not recommended. The best source for consolidated turnover of an enterprise group is their annual report.

Enterprise group – country of global decision centre (non-core)

Purpose

To allocate the enterprise group to the country where decisions on its global strategy are actually taken. Foreign affiliate statistics use the resident country of the global group head.

Definition

Country where the strategic decisions referring to an enterprise group are taken.

Proxies

In most cases, the global decision centre is the global group head. In these cases, the country of global decision centre is the country where the global group head is located. In special cases, the global group head is not able to act as a decision centre for the whole group. Units without substantial physical presence, significant economic activities, or employment are not able to be engaged in decisions. In these cases, the global decision centre should be appointed at the next level downwards in the enterprise group tree structure, where the global decisions of the enterprise group are taken.

Sources

Administrative and commercial sources; statistical surveys. (In the case of the European Union, the EuroGroups register provides standardized information on the decision centre of the multinational enterprises covered.)

Comments

The country where the decision-making centre is located is often referred to as the "nationality of the group", although nationality can be understood in other ways, for instance to refer to the nationality of the owner. The roles of the group head and decision centre may vary a great deal.

This cannot be taken into account in an operational rule, but can be established during profiling. To gain further knowledge about how the group operates, the group can be further analysed. Factors to be considered include where the main production activity is located, where value added is created, and where research and development (which influences the future strategy of the group) take place.

These characteristics are relevant when allocating the enterprise group to the country where decisions on its global strategy are actually taken. Foreign affiliate statistics use the resident country of the global decision centre for defining the outward FATS population (the resident global decision centre is the reporting unit on its foreign affiliates) and the inward FATS population (all enterprises which have a non-resident global decision centre are surveyed).

Coordination with the national statistical offices of the other countries where the multinational enterprise group is acting is recommended in order to agree on the global decision centre and the global economic characteristics of the group for producing international consistent statistics.

Enterprise group – countries where enterprise group has enterprise(s) or local unit(s) (non-core)

Purpose

The characteristic is essential information for outward FATS and for studying intragroup trade, which also needs the names, addresses and identity numbers of legal units belonging to the group.

Definition

Country codes (according to the §4.13 of the Balance of Payments Vademecum).

Sources

Administrative (consolidated balance sheets) and commercial sources, the Euro-Groups register.

A1.4: Enterprise group – relationships between units

(No characteristics.)

European Union example of relationship

The example concerns the European Union, including member States of EFTA.

- The identity number of an MNE comes from the EGR, a supra-national register.
- The identity number (ID) of a truncated group is derived from the common ID of the MNE and thus given centrally by an EGR procedure.
- The ID should remain the same so long as the group is considered to be continuing. (Continuity is discussed in chapter 4 of these Guidelines and in chapter 22 of the Eurostat BR Recommendation manual.)
- The global group head and/or the main decision centre may be moved from one country to another with the enterprise group. Thus, it is not advisable to give an ID that includes a country code.

National numbers can be used until a system of unique Europe-wide identity numbers is established. From then on, a second identity number may be useful to be able to trace the continuity of all-resident groups if they become multinational.

In principle, this approach could be used by any group of countries that agreed to cooperate in handling MNEs. It would need to be supported by legal rights to exchange information, which would likely require the existence of some kind of integrated economic union. In practice, the only known collection of countries where the approach is being applied is the European Union.

A1.5: Enterprise group – relationship with other registers and update sources

Purpose

There can be two kinds of links:

> ➤ Links with sources of information and data on enterprise groups, mostly either private providers or specific administrative registers or enquiries; several examples show that the concerned administrative enquiries might be managed by statistical offices under specific non-statistical regulations.

> ➤ Links with other countries relating to multinational enterprise group (as exemplified in the following box).

Annex A2: Enterprise characteristics

A2.1: Enterprise – identification characteristics

Enterprise – identity number

Purpose

To identify the unit and to follow its continuity and the demographic events leading to discontinuity.

Definition

Given nationally in the SBR.

As the continuity rules for enterprises should be applied, the identity number should remain the same while the enterprise is considered as continuing.

Sources

Assigned by SBR procedures.

Comments

In the case of a 1:1 relation between the legal unit and enterprise, the same identity number is often used but this is not recommended as, conceptually, they are different units.

Enterprise – name

Enterprise – address

Enterprise – email address and website addresses (non-core)

Purpose

To provide contact information. The website can also be used for statistical purposes.

Definition

Countries can decide what information is required to meet their needs.

Proxies

The enterprise name may be the same as the legal unit name, but it may also differ (i.e. be a trade name) even if the enterprise comprises only one legal unit. An enterprise often uses the name of the main legal unit that operates it.

Sources

Administrative sources, websites, SBR procedures, surveys.

Enterprise – Identification number(s) of legal unit(s) of which the enterprise consists

Purpose

To enable the enterprise to be linked to its constitutive legal unit, or legal units if more than one.

Definition

This is simple for enterprises consisting of one legal unit only. In the case of complex enterprises, the identity numbers of all legal units must be recorded.

Proxies

None.

Sources

Created by SBR procedures.

Comments

It is recommended that the dates when the links are created and (possibly) dissolved in the SBR are recorded as well as the reference dates in the real world.

A2.2: Enterprise – demographic characteristics

Enterprise – date of commencement of economic activities in the country

Purpose

The date is required for monitoring the demography of enterprises.

Definition

The date refers to the economic birth of the enterprise.

Proxies

As explained in section 5.3, the date of birth is, in principle, the date on which the first financial commitments are made. In practice it may refer to the registration date in the administrative source if the unit starts its economic activities immediately after that.

Sources

Administrative sources, surveys.

Comments

It is necessary to register the relation with legal units.

The legal unit(s) associated with an enterprise may change and be reregistered, for example after a change of legal form, while the enterprise continues, i.e. remains the same.

Enterprise – date of permanent cessation of activities in the country

Purpose

The date is required for monitoring the demography of enterprises.

Definition

The date refers to the date when an enterprise becomes permanently inactive or dead.

Proxies

As explained for the corresponding characteristic of legal units, this date may not be available with any precision. Only the fact that the enterprise has ceased to exist or has ceased its activities during the reference year may be known, in which case the date should be estimated.

Sources

Surveys, administrative sources or defined by SBR procedures.

Comments

The relationship to legal units needs to be registered.

A2.3: Enterprise – stratification characteristics[69]

[69] Enterprises are defined globally (which means at world level when adequate). For establishing national level, they may be truncated to the national territory of each relevant country. In the case of truncated enterprise, double coding of some characteristics (in particular, principal activity code) at national and supra-national level is recommended.

Enterprise – principal activity code at ISIC four-digit level

Purpose

The principal activity code is a key stratification characteristic. The code is also very important in judging the role of certain units in the enterprise group structure, mainly the group head, the decision centre and SPEs, and in general in enterprise group structuring and demography. The enterprise group's worldwide dispersion of employment by activity can also be studied through its constituent units.

Definition

The activity code is derived according to the rules in ISIC Rev. 4.

Proxies

Other criteria can be used to define the principal activity code, if value added is not available. Employment (preferably) and turnover (with the drawback that it depends on from where it is measured) broken down by the different activities can be used, also activity descriptions.

Sources

Surveys and administrative sources.

Comments

If the enterprise has a legal capacity to operate in different trades, data from administrative sources may indicate a range of activities, but quantitative information on their relative importance may not be available. In such a case the data have to be obtained by survey.

If the enterprise is linked to just one local unit, the principal activity for both types of unit should be the same.

Enterprise – conditional: secondary activities, if any, at ISIC four-digit level

Purpose

Knowledge of secondary activities is very important for large enterprises, for example for short-term business surveys and for national account purposes.

Definition

The activity codes are derived according to the rules in the ISIC Rev. 4.

Proxies

Several criteria can be used for defining the activity code(s), if value added is not available. Employment and turnover by different activities can be used (when available), also activity descriptions.

Sources

Surveys (more accurately) or administrative sources.

Comments

The values of this characteristic may lead to delineation of establishments belonging to the enterprises.

Enterprise – number of persons employed

Enterprise – number of employees

Enterprise – number of employees in FTE (non-core)

(These characteristics are calculated in accordance with the same rules as for local units, establishments or KAUs.)

Purpose

13.1.   Many countries are interested in breakdowns by size class, as small and medium-sized enterprises (SMEs) are often considered to be a major source of new employment. The precise definition of SME has to be decided by each country.[70] The share of SMEs and, within SMEs, of small and micro enterprises can be calculated on the basis of this characteristic.

Employment within the enterprises that make up an enterprise group may be used to determine the continuity or otherwise of the enterprise group.

An enterprise group's worldwide dispersion of employment by activity, and the impact of offshoring, can also be studied by through the employment of its constituent enterprises.

Definition

Number of persons employed: this means the total number of persons who work in the unit, including wage earners and self-employed persons (i.e. working proprietors, partners working regularly in the unit and unpaid family workers), as well as persons who work outside the unit but who belong to it and are paid by it (e.g. sales representatives, delivery personnel, repair and maintenance teams). It excludes workforce members supplied to the unit by other enterprises, persons carrying out repair and maintenance work in the unit on behalf of other enterprises, and those on compulsory military service.

Number of employees: this is defined as those persons who work for an employer and who have a contract of employment and receive compensation in the form of wages, salaries, fees, gratuities, piecework pay or remuneration in kind. A worker from an employment agency is considered an employee of the employment agency and not of the units in which he or she (temporarily) works.

Sources

Surveys or administrative sources. Social security sources may be used as sources of information.

Comments

As an enterprise consists of one or more local units, the employment of an enterprise as of a given date should equal the sum of the employment for the local units. The same holds true for establishments and KAUs belonging to the enterprise.

SBR employment values should be compared with those of other sources (structural business statistics, business demography, labour-force survey) for consistency. Of course, the values may differ if different concepts, definitions or methodologies are used.

[70] For example, in the European Union, according to Commission Recommendation 2003/361/EC, they consist of independent enterprises below 250 persons employed full-time.

Enterprise – turnover

Purpose

For some surveys, mainly repeated ones, and for some spheres of activity, it may not be appropriate to stratify according to employment. Moreover, for accurate calculation purposes, the size of enterprises should also be measured in terms of their turnover.

Turnover is not an additive measure which restricts its direct (non-consolidated) use for the enterprises in the enterprise group framework. However, it can be used for group-related research and development intensity calculations.

Definition

Turnover consists of all revenues arising during the specified reference period, in the course of ordinary activities of the statistical unit and is presented net of all price reductions, discounts and rebates granted by it.

The revenue referred to is arising from contracts with customers and are realized through the satisfaction by the statistical unit of the contractual performance obligations usually represented by the sale (transfer) of goods or the rendering of services, however, it can also contain revenues obtained as a yield on the use by others of the statistical unit's assets.

Excluded from net turnover are:

> All taxes, duties or levies linked directly to the output.

> Any amounts collected on behalf of any principal, if the statistical unit is acting as an agent in its relationship with said principal; turnover includes only the service charge for acting as an agent.

> All income not arising in the course of ordinary activities of the statistical unit. Usually, these types of income are classified as "Other income", "Financial income", "Extraordinary income" or under a similar heading, depending on the respective set of generally accepted accounting standards used to prepare the financial statements.

Proxies

Fiscal sources may be used as sources of information, for example, corporate income tax or VAT returns. There may be cases where such information is unavailable, for example where units are not subject to VAT. One possibility in such a situation is to estimate turnover based on employment using a standard turnover per head ratio calculated by the ISIC class.

Turnover is relevant for market units; it should not be estimated for non-market units. Non-market units, however, can have secondary production.

Sources

Surveys and administrative sources.

Comments

This turnover is "net" in the sense that it excludes the price reductions, discounts and rebates.

The figure used should be the actual turnover. Turnover may be available only with some delay, especially for newly created enterprises. This may cause problems, mainly for business demography, and it may be necessary to estimate the turnover in such cases.

As turnover from different sources is not always the same, using different sources may lead to inconsistencies. Turnover is usually measured excluding VAT.

Turnover is an essential piece of accounting data. Even if comparing turnover between enterprises across different ISIC sections may not make much sense, it may be the only parameter that permits a breakdown by product or allows exports to be ascertained.

Enterprise – institutional sector according to 2008 SNA

Purpose

The institutional sector classification in SBR has several important functions, as noted in chapters 3 and 4.

Definition

The classification of institutional sector applies to the institutional unit which includes also the enterprise.

The classification is detailed in annex B.

Proxies

From the information available, it may not always be possible to define the sector or subsector accurately. In such cases, large and important units could be checked manually. A proxy can be derived from the ISIC code, legal form and country of global decision centre.

Sources

Defined by SBR procedures in cooperation with national accounts, according to established rules, based on SBR information and using administrative sources when necessary.

Comments

In the vast majority of cases, it is possible to calculate the institutional sector code from other explicitly recorded characteristics, especially legal form, ISIC code and controlling country. Algorithms can be used for this purpose. In certain cases the institutional sector must not conflict with the principal activity and whether the enterprise is foreign controlled must be consistent with the respective characteristics of the enterprise group to which it belongs.

The principal activity of the enterprise separates the financial and non-financial corporate enterprises. Two criteria can be applied to separate corporate and quasi-corporate enterprises from enterprises attached to the household sector: the legal form of the legal unit operating the enterprise and, if the legal unit is a natural person, whether or not it keeps annual accounts for the assessment of its tax liability. It may therefore be appropriate to provide an intermediate code indicating, in the case of enterprises operated by a natural person, if they keep accounts or not. That code might perhaps record whether the enterprise is taxed on profits on the basis of its accounts (actual profits) or on the basis of an assessment.

Business statistics cover only market activities and a code indicating whether the activity of an enterprise is market or non-market is essential for them. The distinction is sometimes difficult to make and can be definitively determined only by looking at the accounts.

Public undertaking is defined as any undertaking over which the public authorities may exercise directly or indirectly a dominant influence by virtue of their ownership of it, their financial participation therein, or the rules that govern it. A dominant influence on the part of the public authorities shall be presumed when these authorities, directly or indirectly:

> ‣ Hold the major part of the undertaking's subscribed capital; or
>
> ‣ Control the majority of the votes attaching to shares issued by the undertakings; or
>
> ‣ Can appoint more than half of the members of the undertaking's administrative, managerial or supervisory body.

This could be one legal form or it could be obtained from an administrative source. It may be interpreted in different ways in different administrations. The key issue is that public authorities control public undertakings. The definition does not clearly specify whether public undertaking should be market-oriented and distinguishing between public undertaking and government unit is sometimes difficult and may require the information from its accounts (whether more or less than 50 per cent of production costs are met by sales).

The European System of Accounts (ESA 2010) further divides the institutional sectors of the 2008 SNA into subsectors, with the exception of the institutional sector for NPISHs.

A2.4: Enterprise – relationships between units

Enterprise – identification number of the all-resident or truncated enterprise group to which the enterprise belongs

Purpose

To identify the enterprises belonging to an enterprise group, for example when looking at size classes, in order to avoid confusion between independent enterprises and enterprises belonging to enterprise groups.

It is important also for ensuring confidentiality treatment of the enterprise, for example when all enterprises in a cell belong to the same enterprise group.

Sources

SBR procedures

Comments

The date when enterprise becomes part of the group should also be recorded.

Enterprise – relationship with other registers and update sources

Purpose

(See annex A1.5.)

Annex A3: Establishment characteristics

An *establishment* is also known as a *LKAU*.

A3.1: Establishment – identification characteristics

Establishment – identification number

Purpose

To identify a unit and to be able to record its continuity.

Definition

It is recommended that the SBR register the identification numbers of establishments in accordance with the advice given in section 11.4.11. The identification number of an

establishment should remain the same while the unit continues (according to the continuity rules defined in chapter 6) even when the enterprise to which it belongs changes.

Proxies

Using a unique administrative identity number (if it exists) is also possible, but may cause problems with continuity and thus is not recommended.

Sources

SBR procedure, unless an administrative number is used.

Establishment – name

Establishment – address

Establishment – telephone and fax numbers, email address and information to permit electronic collection of data (non-core)

Purpose

To contact the unit.

Definition

The official name of an establishment is generally the same as the legal unit that controls it, with some additional part (usually) specifying location or activity. If there is only one establishment belonging to the legal unit, a separate name may not exist.

The actual address of the location of the unit and the contact information should always be recorded.

Proxies

Different establishments within a legal unit may use different trading names (also known as "signboard names" or "commercial names") and provision should be made for recording these different trading names where they exist.

Sources

Administrative sources and surveys.

Comments

The address could be given a code referring to a national territorial nomenclature. It would be preferable if that nomenclature or national geographical code enables the unit to be pinpointed as accurately as possible, at the level of the street, section of street and building.

In addition to the physical address of the establishment it is useful to provide for the possibility of recording a correspondence address where different.

Establishment – identity number of the enterprise to which it belongs

Purpose

The establishment must be linked with the enterprise to which it belongs.

Definition

This link can be included in the register by adding the identity number of the enterprise to the establishment file (and vice versa).

Proxies

Other ways are conceivable, for example when the enterprise consists of one establishment only, another (simple) arrangement of the business register is possible.

Sources

SBR procedures.

A3.2: Establishment – demographic characteristics

Establishment – date of commencement of activities (economic birth)

Purpose

The date is required for monitoring the demography of establishment.

Definition

This date should refer to the birth or other creation date of the unit according to the continuity rules.

Establishment – date of final cessation of activities (economic death)

Purpose

The date is required for monitoring the demography of establishments.

Definition

The date refers to the death of a unit.

Proxies

As explained for the corresponding characteristic for legal units, this date may not be available with any precision. Only the fact that the establishment has ceased to exist or has ceased its activities during the reference year may be known, in which case the date should be estimated.

Sources

Survey, administrative source and SBR procedures.

A3.3: Establishment – economic stratification characteristics

Establishment – principal activity code at ISIC four-digit level

Purpose

The principal activity code is a stratification characteristic and important in the compilation of regional and small area statistics.

Definition

The activity code is determined according to the rules in the ISIC Rev. 4 handbook. The activities actually pursued in the unit are taken into account. They may not coincide with the principal activity of the enterprise.

Proxies

Several criteria can be used for defining the principal activity code. Employment by different activities (if available) can be used, also activity descriptions.

Sources

The code can be obtained from surveys or administrative sources.

Comments

If the activities are ancillary in the context of the enterprise, the value recorded should indicate this.

Establishment – secondary activities, if any, at ISIC four-digit level (non-core)

Purpose

Secondary activities may be used in determining establishments that are of key importance for national/regional accounts.

Definition

Activity codes are allocated according to ISIC Rev 4.

Proxies

Several criteria can be used for defining secondary activity code(s). Employment by different activities can be used (if available), but also activity descriptions, etc.

Sources

Surveys and administrative sources.

Establishment – activity carried out in the unit constituting an ancillary activity of the enterprise to which it belongs (yes/no) (non-core)

Purpose

To distinguish ancillary units. This characteristic enables statistical analyses to reallocate the cost of ancillary activities to the activities for the benefit of which they are undertaken.

Definition

An ancillary activity is incidental to the main activity of an enterprise. It facilitates the efficient running of the enterprise but does not normally result in goods and services that can be marketed. For enterprises that are relatively small and have only a single location, ancillary activities are not separately identified. For larger enterprises with multiple locations, it may be useful to treat ancillary activities in the same way as a secondary or even a principal product (2008 SNA, para. 5.10).

Proxies

Activities in certain ISIC classes often constitute ancillary activities.

Sources

Administrative sources and surveys.

Comments

If a unit has been identified as an ancillary unit, this fact should be positively indicated in the SBR.

Establishment – number of persons employed

Establishment – number of employees

Purpose

The SBR should record the actual numbers of persons employed and employees, the latter both as a head count and as an FTE.

The main aim is to provide a stratification characteristic. Persons employed are preferable for stratifying survey samples for very small units.

In addition, SBR employment figures are sometimes disseminated directly, especially as small area statistics where the SBR may be the only comprehensive source.

Definition

For stratification purposes, the SBR aim is to provide end of year values (including seasonally active units). As the end date approach is not harmonized across countries, the annual average can also be used.

The number of employees as an FTE may be calculated for a full year as well as for the period in which an enterprise is active. If the number of employees in FTE is used as stratification characteristic, it should be calculated as an average over the active period,

while in compiling annual statistics it should be calculated as the average over the full 12 months. 0 means less than half a person, whether calculated as a head count or FTE.

Proxies

Number employed can be obtained directly in some countries in administrative sources, while in other countries only an administrative source providing number of *paid employees* is available. However, number employed can be calculated from data on paid employees by adding working proprietors and unpaid employment calculated according to legal form and activity. In particular:

- For sole proprietors, total employment = paid employees + 1
- For partnerships, total employment = paid employees + number of partners

Depending on the availability of administrative sources, more sophisticated methods have been devised in some countries.

Sources

Administrative sources, surveys, calculations.

Comments

The reference period used for the measurement of employment in business demography is a year; in other words, the labour force should be an annual average, although this can be approximated by using the number of persons employed at any given moment during the year if this is the only information available.

Both head counts and FTEs have certain advantages and the latter should be recorded if possible. Head count is the number of physical persons, full-time and part-time, employed by a unit. FTE is the number of FTE jobs, defined as total hours actually worked divided by average number of hours actually worked in full-time jobs. FTEs are a more accurate measure of labour input but they are more difficult to measure as additional data would be required which is difficult to survey. As the concept of "full-time" may vary across economic sectors and countries, the definition does not really make the FTE data comparable. Given the administrative origin of the data, it may not be possible to calculate FTEs in some countries. Another possibility would be to use "hours actually worked" directly. This is gaining favour in employment statistics, but the comment on data availability is also likely to apply to this characteristic.

Establishment – geographical location code (non-core)

Purpose

The geographical location code complements the address and postal codes and can be used to derive classifications relating to the geographical location of units at the most detailed level, in addition to other national classifications such as administrative regions, travel-to-work areas, health regions, and education regions.

Definition

Countries can decide which code is most useful for their own purpose.

Proxies

The geographical location code can refer to a classification at the most detailed level used in the country. It can refer to geocoding according to latitude and longitude points.

Sources

Administrative sources.

A3.4: Establishment – relationship between units

Establishment – identification number of the enterprise to which the establishment belongs

Purpose

To identify the establishments belonging to an enterprise: for example to link geographical details to the economic enterprise data.

Sources

SBR procedures.

Annex A4: Local unit of enterprise characteristics

A4.1: Local unit of enterprise – identification characteristics

Local unit of enterprise – identification number

Purpose

To identify the unit and to be able to record its continuity.

Definition

The identity number of a local unit should remain the same while the unit continues according to the continuity rules defined in chapter 6 (even when the legal unit to which it belongs changes).

Proxies

Using a unique administrative identity number (if it exists) is also possible, but it may cause problems with continuity and thus is not recommended.

Sources

SBR procedure, unless the administrative number is used.

Local unit of enterprise – name

Local unit of enterprise – address

Local unit of enterprise – telephone and fax numbers, email address and information to permit electronic collection of data (non-core)

Purpose

To contact the unit.

Definition

The official name of the local unit of an enterprise (and truncated enterprise) is generally the same as the legal unit that controls it, with some additional part usually specifying location or activity. If there is only one local unit in the legal unit, a separate name may not exist.

The actual address of the location of the unit and the contact information should always be recorded.

Proxies

Different local units within a legal unit may use different trading names, also known as "signboard names" or "commercial names", and provision should be made for recording these different trading names where they exist.

Sources

Administrative sources and surveys.

Comments

The address could be given a code referring to a national territorial nomenclature. It would be preferable if that nomenclature or national geographical code enables the unit to be pinpointed as accurately as possible, at the level of the street, section of street and building.

It is always useful to provide for the possibility of recording a correspondence address in addition to the physical address of the local unit if the unit wishes correspondence relating to statistical surveys to be sent elsewhere.

Local unit of enterprise – identity number of the enterprise to which it belongs

Purpose

The local unit of enterprise (and truncated enterprise) must be linked with the enterprise to which it belongs.

Definition

This link can be included in the SBR by adding the identification number of the enterprise to the local unit of enterprise (and truncated enterprise) (and vice versa).

Sources

SBR procedures.

A4.2: Local unit of enterprise – demographic characteristics

Local unit of enterprise – Date of commencement of activities (economic birth)

Purpose

The date is required for monitoring the demography of local units.

Definition

This date should refer to the birth or other creation date of the unit according to the continuity rules.

Sources

SBR procedures.

Local unit of enterprise – date of permanent cessation of activities (economic death)

Purpose

The date is required for monitoring the demography of local units.

Definition

The date refers to the death of the unit or the date on which it became permanently inactive.

Proxies

As explained in connection with the corresponding characteristic of legal units, this date may not be available with any precision. Only the fact that the local unit has ceased to exist or has ceased its activities during the reference year may be known, in which case the date should be estimated.

Sources

Surveys, administrative sources, SBR procedures or estimated.

A4.3: Local unit of enterprise – economic stratification characteristics

Local unit of enterprise – principal activity code at ISIC four-digit level

Purpose

The principal activity code is a stratification characteristic and is important in the compilation of regional and small area statistics.

Definition

The activity code is determined according to ISIC Rev 4. The activities actually conducted by the unit are what matters, even if they are ancillary in the context of the enterprise.

Proxies

Several criteria can be used in defining principal activity code. Employment by each different activity can be used (if available), also activity descriptions.

Sources

Surveys and administrative sources.

Comments

If the local unit activity is ancillary in the context of the enterprise, this should be indicated.

Local unit of enterprise – secondary activities, if any, at ISIC four-digit level (non-core)

This concerns only local units which are the subject of surveys

Purpose

Secondary activities may be used for helping determining establishments, which are of key importance for national and regional accounts. Establishments can be recorded in the SBR as separate units (this practice is commonly applied by many smaller countries). If this is not the case, this characteristic offers the possibility of identifying them for analytical purposes.

Definition

The activity codes are allocated in accordance with ISIC Rev. 4.

Proxies

Several criteria can be used for defining the secondary activity code(s). Employment by each different activity can be used (if available), also activity descriptions.

Sources

Surveys or administrative sources.

Local unit of enterprise – Activity carried out in the unit constituting an ancillary activity of the enterprise to which it belongs (yes/no) (non-core)

Purpose

To distinguish ancillary units. This characteristic enables statistical analyses to reallocate the cost of ancillary activities to the activities for the benefit of which they are pursued.

Definition

An ancillary activity is incidental to the main activity of an enterprise. It facilitates the efficient running of the enterprise but does not normally result in goods and services that can be marketed. For enterprises that are relatively small and have only a single

location, ancillary activities are not separately identified. For larger enterprises with multiple locations, it may be useful to treat ancillary activities in the same way as a secondary or even a principal product (2008 SNA, para. 5.10).

Proxies

Certain ISIC classes often constitute ancillary activities.

Sources

Administrative sources, surveys.

Comments

Whether or not a unit is an ancillary unit should be explicitly recorded in the SBR. The absence of a value should be interpreted as "having no knowledge about this characteristic".

Local unit of enterprise – number of persons employed

Local unit of enterprise – number of employees

Local unit of enterprise – number of employees in FTE (non-core)

Purpose

The SBR should record the actual numbers of persons employed and employees, both as head counts and the latter also in FTEs. The main aim is to obtain stratification characteristics as well as statistical information. Persons employed are preferable for stratifying survey samples for very small units. The business register employment figures are used especially for small area statistics, where the business register is the only comprehensive source. In addition, employment figures can be used by employment statistics where needed.

Definition

The structural business statistics definitions should be used, with the exception that the requirement to measure the numbers as annual averages does not apply. For stratification purposes and in accordance with the intention of the SBR the situation at the end of year (including seasonally active units) may be used. As the end date approach is not harmonized, the annual average can also be used as reference calculated for a certain period. The number of employees in FTEs might be calculated for a full year as well as for the period in which the enterprise is active. If the number of employees in FTEs is used as a stratification characteristic, the calculation for the active period should be used, while for statistics covering a year the calculation should cover the whole year.

Proxies

These figures can be obtained directly in some countries, while other countries may have an administrative source available only for the number of paid employees. The latter countries can, however, obtain total employment by making a statistical adjustment to their figures on paid employees by adding a constant representing working proprietors and any other form of unpaid employment calculated according to legal form and activity, for example:

> ▸   For sole proprietors, total employment = paid employees + 1
>
> ▸   For partnerships, total employment = paid employees + number of partners

Depending on the availability of administrative sources, more sophisticated methods have been devised in some countries.

Sources

Administrative sources, surveys, calculations.

Comments

Note that the reference period used for the measurement of employment in business demography is a year; in other words, the labour force should be an annual average, although this can be approximated by using the number of persons employed at any given moment during the year if this is the only information available. How the annual average is calculated depends on the updating frequency of the register. If the unit operates during only part of the year (seasonal, new enterprises), the average should be calculated for that period.

Both head counts and FTEs have certain advantages and the latter should be recorded if possible. Head count is the number of physical persons, full-time and part-time, employed by a unit. FTE is the number of FTE jobs, defined as total hours actually worked divided by average number of hours actually worked in full-time jobs. FTEs are a more accurate measure of labour input but they are available in fewer countries. As the concept of "full-time" may vary, the definition does not really make the FTE data comparable. Given the administrative origin of the data, it may not be possible to calculate FTEs in some countries. Another possibility would be to use "hours actually worked" directly. This is gaining favour in employment statistics, but the comment on data availability is also likely to apply to this characteristic.

Local unit of enterprise – geographical location code (non-core)

Purpose

The geographical location code complements the address and postal codes and can be used to derive classifications relating to the geographical location of units at the most detailed level. Other national classifications such as administrative regions, travel-to-work areas, health or education regions, etc. can also be derived from it.

Definition

Member States can decide which code is most useful for their own purpose.

Proxies

The geographical location code may refer to classification at the most detailed level used in the country. It may refer to geocoding to latitude and longitude points recorded by GPS in countries where the exact site of local unit can be recorded.

Sources

Administrative sources.

A4.4: Local unit of enterprise – relationship between units

Local unit of enterprise – identification number of the enterprise to which the local unit belongs

Purpose

To identify the local units belonging to an enterprise: for example to link geographical details to the economic enterprise data.

Sources

SBR procedures.

Annex A5: Legal unit characteristics

A5.1: Legal unit – identification characteristics

Legal unit – identity number

Purpose

To identify the unit and to link it with other units in the SBR and with administrative and statistical sources.

Definition

The identity number of the legal unit can be either specific to the SBR or set by an administrative source and used by the SBR, or shared by several administrative sources and used by the SBR.

Sources

If the identity number of the legal unit assigned by an administrative source is used by the SBR, then its updating has to follow the changes taking place in the administrative source, even if it is an issue for economic purposes. For instance, a change of legal form from natural to legal person may result in the fiscal administration assigning a different identifier to the same economic unit.

If the identity number of the legal unit in the SBR is not the administrative identifier, the administrative identifier should be handled as a characteristic and a record of its changes should be kept.[71]

Comments

Common business identifiers, shared with fiscal and other government departments, greatly facilitate the connection of the SBR with other administrative sources.

The continuity rules for legal units depend on national legislation. There are no general recommendations.

Legal unit – name

Legal unit – address

Legal unit – telephone and fax numbers, email address and information to permit electronic collection of data (non-core)

Purpose

To contact the unit.

In the event that a common identifier is missing, names and addresses can also be used for linking units.

Definition

The addresses should be recorded at the most detailed level possible and respect international standards.

The information may refer to a legal or a natural person. In the case of a legal person, the official business name and address must be recorded. In the case of a natural person, it may be useful to maintain both a business and a personal address. In addition, the following information may be recorded:

- ▸ Family name(s)
- ▸ Names normally used and possible pseudonyms
- ▸ First names
- ▸ Gender (for address purposes — Dear Mr./Ms. ...)

[71] This could be the situation if an administrative legal unit is given a unique identifier related to the region in which it is registered. If the unit moves to another region within the country and then is given a new identifier, there will not actually be a unique administrative identifier for the unit for that country.

This information is often insufficient to identify a natural person with certainty. It may therefore be useful also to record either the date of birth or the personal identity number.

Proxies

In some countries and for some legal forms, the business name may be very long and have to be abbreviated. In this case, strict rules on abbreviation must be established and applied.

Where applicable, territorial classifications and nomenclatures may be used for coding of addresses. It is even better if addresses respect the International Organization for Standardization (ISO) standard.

Sources

Administrative sources, mainly trade and company registers and surveys.

Comments

Legal persons and also sole proprietors often use initials, an acronym or a trading name instead of their official name in their business or administrative relations. There must be provision for recording this information separately.

A5.2: Legal unit – demographic characteristics

Legal unit – date of incorporation for legal person or date of official recognition as an economic operator for natural person

Purpose

The characteristic is needed for the inclusion of new units.

Definition

The "date of official recognition" should be the date on which an identification number is given, or the date on which the legal existence was approved, be it via a company or trade register, a VAT register or other register.

Proxies

If the exact date is not available, the year from which the unit has been monitored can be used as proxy, including a certain date (like 1 January) that is indicative of the situation, as defined in SBR procedures.

Sources

Administrative sources: trade or company register, tax administration, social security.

Comments

In general, the date is the prerequisite for a unit to engage in legal economic transactions. Given that the SBR is usually supplied from administrative sources, a date for official recognition should always exist and be stored in the SBR. The legal unit may start its economic activity (and only then be regarded as an enterprise or part thereof) with some delay after its recognition, or it may remain economically inactive. If a legal unit remains inactive, it can either be omitted from the register, or kept in the SBR but marked as inactive. The latter approach is preferable. Keeping legally alive but economically inactive legal units in the SBR facilitates the use of data from administrative sources.

Legal unit – date of commencement of economic activity

Purpose

To ensure that a legal unit that has started economic production is recognized as an active enterprise.

Definition

The "date of effective economic activity" should be the date on which the unit starts its economic activity. This date may be declared by the unit when it registers officially and receives an identification number, or when its legal existence was approved, be it via a company or trade register, a VAT register or another register.

Proxies

If the exact date is not available, the year from which the unit has begun its activities can be used as proxy, including a certain date (like 1 January) defined by the SBR procedures.

Sources

Administrative sources: trade or company register, tax administration, social security.

Comments

In general, the date is the prerequisite for a legal unit to engage in legal economic transactions. Given that the SBR is usually supplied from administrative sources, a date for an effective economic activity should always exist and be stored in the SBR.

Legal unit – date of permanent cessation of activities (economic death)

Legal unit – date on which the legal unit ceased (administrative death)

Purpose

These demographic characteristics are needed for monitoring the death or permanent inactivity of the unit from both economic and administrative perspectives.

Definition

A legal unit ceases to be part of an enterprise (economic death) when:

> ‣ The unit ceases to be economically active and it is not part of the control chain within the enterprise group.

> ‣ The unit ceases to exist.

The record of a legal unit that has ceased (administrative death) should be kept in the SBR marked "dead".

Proxies

Registration of the year of economic death is important, even if the precise day and month are not known, as is often the case.

Sources

Administrative sources (from which the date often comes with considerable delay), surveys.

Comments

A legal unit usually has no particular interest in officially announcing cessation of its activities. Cessation may be the end result of a slow process of diminishing activity. Even if the activity completely ceases, the owner may still want to retain the option of resuming in the future and hence may retain a legal name, fiscal number and other legal attributes.

Between economic death and administrative death there is often a period of inactivity during which the unit may be regarded as economically inactive (also called "dormant"). Signs of this situation are lack of employees, the cessation of tax compliance and inability to contact the unit after repeated efforts.

A5.3: Legal unit – economic stratification characteristics

Legal unit – legal form

Purpose

Legal form (also known as "legal status") is useful in helping to eliminate ambiguity in identification searches. It is also a possible criterion for selection of units into survey frames and for stratification for survey sampling. It is also useful in defining the institutional sector (see annex B). In monitoring the internal market there is an interest in being able to distinguish publicly traded and incorporated companies.

Definition

Defined according to national legislation. Section 4.6.4 gives an example for European Union countries.

Sources

Administrative sources, surveys. The legal form of a legal or natural person determines the tax regime applicable to the unit. Thus, if the SBR is supplied with taxation records it has access to the values of this characteristic.

Comments

Adjustments to questionnaires and collection processes are often made in accordance with legal form of the legal unit operating an enterprise.

Legal unit – principal activity

Purpose

As a stratification characteristic and for the compilation of regional and small area statistics.

Definition

The activity code is determined according to ISIC Rev. 4. The code is based on the activities actually conducted by a legal unit even when they are ancillary in the context of the enterprise.

Proxies

Employment or value added for each of the different activities can be used (if available), also activity descriptions and others.

Sources

Surveys and administrative sources.

Comments

This characteristic indicates whether the activities are ancillary in the context of the enterprise.

Legal unit – secondary activities (non-core)

Purpose

For indirect use by national/regional accounts to estimate enterprise or establishment figures.

Definition

The activity codes are determined according to ISIC Rev. 4. The code is based on the activities actually conducted by a legal unit even when they are ancillary in the context of the enterprise.

Proxies

Employment or value added for each of the different activities can be used (if available), also activity descriptions and others.

Sources

Surveys and administrative sources.

Legal unit – ancillary unit (yes/no) (non-core)

Purpose

To distinguish ancillary legal units. This characteristic enables statistical analyses to reallocate the cost of ancillary activities to the activities for the benefit of which they are conducted.

Definition

An ancillary activity is incidental to the main activity of a unit. It facilitates the efficient running of the unit but does not normally result in goods and services that can be marketed. For units that are relatively small and have only a single location, ancillary activities are not separately identified. For larger units with multiple locations, it may be useful to treat ancillary activities in the same way as a secondary or even a principal product (2008 SNA, para. 5.10).

Proxies

Ancillary activities tend to relate to certain ISIC classes, for example bookkeeping. While bookkeeping is a potential ancillary activity, however, it is predominantly performed by the businesses that offer bookkeeping services on the market and constitute the corresponding ISIC class.

Sources

Surveys and administrative sources.

Comments

Whether or not a unit is an ancillary unit should be explicitly recorded in the SBR. The absence of a value should be interpreted as "having no knowledge about this characteristic".

Legal unit – turnover (with flag for consolidated turnover)

Purpose

As a size indicator for stratification, and for profiling an enterprise group using a bottom-up approach.

Definition

Turnover consists of all income arising during the reference period, in the course of ordinary activities of the unit, and is presented net of all price reductions, discounts and rebates granted by it.

The income referred to arises from contracts with customers and is realized through the satisfaction by the unit of the contractual performance obligations usually represented by the sale (transfer) of goods or the rendering of services. It can also contain revenues obtained as a yield on the use by others of the statistical unit's assets, however.

Excluded from net turnover are:

>   ▸   All taxes, duties or levies linked directly to revenue.

> ➤ Any amounts collected on behalf of any principal, if the statistical unit is acting as an agent in its relationship with said principal.

> ➤ All income not arising in the course of ordinary activities of the statistical unit. Usually, these types of income are classified as "Other income", "Financial income", "Extraordinary income" or under a similar heading, depending on the respective set of generally accepted accounting standards used to prepare the financial statements.

Proxies

Fiscal sources may be used as sources of information, for example, corporate income tax or VAT returns. There may be cases where such information is unavailable, for example where units are not subject to VAT. In such a situation one possibility is to estimate turnover based on employment using a standard turnover per head ratio calculated by the ISIC class.

Turnover is relevant for market units; it should not be estimated for non-market units. Non-market units can have secondary production, however.

Sources

Administrative sources and surveys.

Legal unit – number of persons employed

Legal unit – number of employees

Legal unit – number employees in FTE (non-core)

Purpose

The SBR should record the actual numbers of persons employed and employees, both as head counts and the latter also in FTEs. The main aim is to obtain stratification characteristics as well as statistical information. Persons employed are preferable for stratifying survey samples for very small units. The SBR employment figures are used in particular for small area statistics, where the SBR is the only comprehensive source. In addition, employment figures can be used by employment statistics where needed.

Definition

Number of persons employed is defined as the total number of persons who work in the unit, including wage earners and self-employed persons, namely, working proprietors, partners working regularly in the unit and unpaid family workers, and also persons who work outside the unit but who belong to it and are paid by it, such as sales representatives, delivery personnel and repair and maintenance teams. It excludes labour supplied to the unit by other enterprises, persons carrying out repair and maintenance work in the unit on behalf of other enterprises and those on compulsory military service.

Number of employees is defined as those persons who work for an employer and who have a contract of employment and receive compensation in the form of wages, salaries, fees, gratuities, piecework pay or remuneration in kind. A worker from an employment agency is considered to be an employee of that employment agency and not of the units in which they (temporarily) work.

For stratification purposes the intention is to use the situation at the end of year (including seasonally active units). As the end date approach is not harmonized, the annual average can also be used as reference calculated for a certain period. The number of employees in FTEs might be calculated for a full year, and also for the period in which the legal unit is active. If number of employees in FTEs is used as a stratification

characteristic, the calculation for the active period should be used, while for statistics covering a year the calculation should cover the whole year.

Proxies

These figures can be obtained directly in some countries, while other countries may have an administrative source available only for the number of paid employees. The latter countries can, however, obtain total employment by making a statistical adjustment to their figures on paid employees by adding a constant representing unpaid employment (including working proprietors), calculated, for example, according to legal form and activity:

- ‣ For sole proprietors, total employment = paid employees + 1
- ‣ For partnerships, total employment = paid employees + number of partners

Depending on the availability of administrative sources, more sophisticated methods have been devised in some countries. Note also that 0 means less than half a person, whether calculated as head counts or FTEs.

Sources

Administrative sources, surveys, calculations.

Comments

Note that the reference period used for the measurement of employment in business demography is a year; in other words, the labour force should be an annual average, although this can be approximated by using the number of persons employed at any given moment during the year if this is the only information available. How the annual average is calculated depends on the updating frequency of the register. If the unit operates during only part of the year (seasonal, new enterprises) the average should be calculated for that period.

Both head counts and FTEs have certain advantages and the latter should be recorded if possible. Head count is the number of physical persons, full-time and part-time, employed by a unit. FTEs are defined in structural business statistics (SBS) (variable 16 14 0) and also in national accounts (FTE employment is the number of FTE jobs, defined as total hours actually worked divided by average number of hours actually worked in full-time jobs). FTEs are a more accurate measure of labour input but they are available in fewer countries. As the concept of "full-time" may vary, the definition does not really make the FTE data comparable. Given the administrative origin of the data, it may not be possible to calculate FTEs in some countries. Another possibility would be to use "hours actually worked" directly. This is gaining favour in employment statistics, but the comment on data availability is also likely to apply to this variable.

Legal unit – institutional sector and subsector in the European System of Accounts (ESA 2010)

Purpose

The institutional sector classification for legal unit is determined by the institutional sector classification of the enterprise with which it is associated.

Sources

SBR procedures

Comments

In practice, institutional sector classification is often applied directly to legal units. Then the enterprise inherits its institutional sector classification from the legal unit(s) from which it is constituted.

In practice too, in the majority of cases, enterprise and legal unit are the same.

A5.4: Legal unit – relationship between units

Link with enterprise group and truncated group

Legal unit – percentage of control of legal units by the enterprise group (direct + indirect)

Purpose
To allow the delineation of the enterprise group.

Sources
SBR procedures.

Legal unit – percentage of ownership of legal units (direct + indirect)

Purpose
To support calculation of control.

Sources
SBR procedures.

Link between enterprise and legal unit

Legal unit – identification number of the enterprise(s) to which the legal unit belongs

Purpose
To identify the legal units belonging to an enterprise.

Sources
SBR procedures

Links with local units

Legal unit – Identification number of the local unit(s) that belongs to the legal unit

Purpose
To identify the local units belonging to the legal unit.

Sources
SBR procedures.

Consolidation method (non-core)
Integration method of the legal unit in the consolidated accounts (if relevant).

A5.5: Legal unit – Relationship with other registers/update sources

Legal unit – Value added tax (VAT) identification number

Legal unit – Identification number of each other administrative source used to maintain the SBR or to compile economic statistics

Purpose

For linking administrative data to the SBR; for the links to non-resident units and for the links to foreign trade.

Definition

Defined by the corresponding administrative source. A separate VAT identification number may not exist in countries where a single identification system for legal units is used by several administrative sources.

Proxies

Certain economic activities may be exempted from VAT and thus have no VAT number. In this case, identity number from another administrative source, such as tax number, corporate registration number may play the same role.

Sources

Mainly tax administration.

Comments

In some countries, a VAT number may relate to part of a legal unit or to a natural person who is not an economic operator but who has a VAT number for tax reasons.

Legal unit – balance sheet data registration identification number

(Applicable only for legal units required to publish accounts.)

Legal unit – balance of payments register or FDI register

Legal unit – farm register

Legal unit – chamber of commerce and trade associations

Purpose

The first link concerns the balance sheet data. Many countries use the published accounts as a source for the SBR and combining the SBR and published accounts data is likely to become very important in the future. It reduces response burden and serves the production of economic and financial statistics.

The second link concerns the balance-of-payments and FDI registers and the usefulness of these links concerns the harmonization of statistics related to globalization. Conventional bank settlements data are becoming more and more frequently replaced by data based on direct surveys, for which the balance-of-payments compilers are increasingly relying on SBRs.

The third link to farm register (if separate from the SBR) is important for the coverage of main agricultural enterprises and for updating the increasing rural multi-activities, where the role of agriculture as principal or secondary activity may often change.

The fourth (set of) link(s) to (often) compulsory registrations that exist in countries for various reasons, for example establishment of legal persons and liability commitments of producers. These administrative or para-administrative bodies may well register births, deaths and changes in legal structures of the legal units within their realms of responsibility. They often cover a large part but not all of the legal units in scope for the SBR.

Definition

The practical arrangement of the links, either from the SBR to the associated registers, or vice versa, is a country matter.

Proxies

The links can be achieved in several ways, for example:

> ‣ Recording in the SBR the reference number of the unit in the other register(s), together with the legal unit identity number.

> ‣ Adding one or more marks to the SBR to indicate that the legal unit is also present in other register(s) under the same identity number.

> ‣ In the absence of a unique identifier, the link can also be built by name and address matching and possible use of other characteristics. This is less effective, but matching names and addresses can be useful for detecting errors. For every legal unit recorded in administrative sources, the different relations should be stored separately.

Sources

Administrative sources.

Comments

The requirement to publish annual accounts depends on national legislation, which may vary between countries. It usually applies to incorporated and publicly traded companies.

The data from published accounts may be considered free from confidentiality restrictions and thus suitable for exchange with other national statistical offices. This is not the case, however, if the data are accompanied by data collected via surveys.

The links to balance sheet data can be used for combining SBR and accounts data, which are generally available in satellite registers.

In the European Union, either the national central bank or the national statistical office is a member of the European Committee of Central Balance Sheet Data Office, which supplies data for an increasing number of member States of the European Union.

The balance of payments register in most countries is maintained by the national central bank. In this case, close cooperation and exchange of information between the national statistical office and the national central bank is vital for well-targeted and good-quality balance-of-payments and FDI surveys. The balance of payments register may alternatively be maintained by the national statistical office, or there may not be a separate balance of payments register and the SBR is used for balance of payments purposes.

The link between farm and business registers is discussed in section 3.4.1.

Several other registers, such as shop registers, tourist establishment registers, transport registers, educational institute registers and others, may also be important for updating the SBR, especially in determining whether or not units are active, because data in these registers tend to be updated frequently.

Harmonization of units in the SBR and these associated registers is an important issue.

Annex B
International classifications

Annex B1: International Standard Industrial Classification of All Economic Activities, Revision 4

For economic statistics, including the SBR, it is recommended that countries apply the International Standard Industrial Classification of All Economic Activities (ISIC), Revision 4. In ISIC, economic activities are subdivided in a hierarchical, four-level structure of mutually exclusive categories. This facilitates data collection, consistent aggregation and presentation and analysis at detailed levels of the economy in an internationally comparable, standardized way. Revision 4 is the most recent version of the standard and is specifically designed to classify enterprises, kind-of-activity units (KAUs) and establishments as defined in the 2008 SNA according to their principal economic activity. ISIC includes four levels of aggregation:

  ‣ *Sections*: The sections are the highest level of aggregation. The sections are labelled alphabetically and listed in figure B1.

  ‣ *Divisions*: The next level contains two-digit divisions. Examples are Division 03: Fishing and aquaculture; Division 13: Manufacture of textiles.

  ‣ *Groups*: The third level contains three-digit groups. Examples are Group 031: Fishing; Group 131: Spinning, weaving and finishing of textiles.

  ‣ *Classes*: The most detailed level contains four-digit classes. Examples are Class 0312: Freshwater fishing; Class 1312: Weaving of textiles.

At each level of ISIC, a statistical unit is assigned to one and only one ISIC code. The set of statistical units that are classified into the same ISIC category is often referred to as an "industry". Examples are:

  ‣ The *furniture industry* – all units classified in ISIC division 31 (Manufacture of furniture).

  ‣ The *construction industry* – all units classified in ISIC section F (Construction).

For national purposes the level of detail provided by ISIC may not be viewed as sufficiently detailed. Countries may thus introduce a fifth hierarchical level, usually called subclass level, or intermediate levels of aggregation (for example between divisions and the much more detailed group level). The national aggregates should, however, be consistent with the ISIC hierarchy. Conversely, it could also be the case that the level of detail in ISIC may be viewed as being too detailed. In such cases, a country might for instance use the three-digit level instead of the four-digit one.

Where national classifications are used these should be fully comparable with ISIC. If this is not the case, international comparability is not ensured. Transformation of

data from one classification to another with different groupings is likely to create problems by leaving some activities out or including the same activity more than once, which reduce data quality and international comparability. It should, therefore, be possible to aggregate the national groupings into the sections, divisions, groups and classes of ISIC. Activities should not be moved between different levels of aggregation or within the same level of aggregation.

Assigning classification codes

ISIC class descriptions and explanatory notes are used in determining the class of activities under which a particular statistical unit is classified. It is recommended that all units are classified down the class level of ISIC, all higher aggregation levels can then easily be derived. Classifying at the class level does not mean that data have to be published at this level. The level of publication has to be decided based on various criteria: confidentiality criteria, type of statistical unit, sample selection, and others.

In order to classify a unit according to the classification rules of ISIC, information on the economic activities performed is needed, together with information on the size of the activities to determine the principal activity. Information on the activities performed might be collected by its being integrated into regular economic surveys, or might be available from administrative sources or from the Internet. It should be noted that the mere name of a company might not always be a good indication of the principal activity.

A unit is to be classified according to its principal (or main) activity. The principal activity of a unit is defined as the activity which contributes most to the total value added of the unit (2008 SNA, para. 5.8). In determining the principal activity, all activities of the unit, except the ancillary activities, are considered but only the principal activity is used to classify the unit. The other activities of the unit are named secondary activities and may also be recorded in the SBR.

Principal and secondary activities cannot be carried out without the support of a number of ancillary activities, such as bookkeeping, transport, storage, purchasing, sales promotion, cleaning, repair and maintenance, security, and others. At least some of these activities can be found in every economic entity. Ancillary activities are not produced for selling on the market, but are only undertaken for the unit itself. Thus, the output of ancillary activities is always intended for intermediate consumption within the same entity. Although most ancillary activities produce services, some goods-producing activities may, as an exception, be regarded as ancillary. The goods produced, however, may not become a physical part of the output of the principal or secondary activities (such as tools).

Top-down method to determine principal activity

A unit may perform one or more economic activities. If all these activities fall into one category in ISIC, the coding is quite straightforward. If, however, these economic activities fall under different ISIC categories, the principal activity has to be determined using the top-down method. The principle of the top-down method is that the classification of a unit at the lowest level of the classification must be consistent with the classification of the unit at higher levels. To satisfy this condition the process starts with the identification of the category at the highest level which comprises most of the value added. The method will then be applied to the second classification level: within the category that accounted for the highest share of value added at the highest level, the category at the second level is determined analogously: it is that category which accounts for the highest share of value added. This method is further applied until the

lowest level is reached. The category at the lowest level determines the principal activity of the unit. It may not necessarily be the category which accounts for more than 50 per cent of the value added.

Input and output-based indicators of value added

In practical implementation the main problem is that usually data on value added by the different activities of the unit is not available and can also not easily be observed. Accordingly, some value added substitutes are proposed in ISIC. They fall into two types: substitutes based on output, such as gross output attributable to the goods and services produced, value of sales of the different goods and services, and substitutes based on input indicators, such as wages and salaries attributable to the different activities, hours actually worked or employment attributable to the different activities. Such substitute indicators should be used as proxies for the unknown value added data to obtain the best approximation possible compared to the results which would have been obtained on the basis of the value added data.

Both the input and output substitutes have problems, however. Generally, the relation of a substitute indicator (such as turnover or employment) to the value added might be quite different for the different activities and thus distort the results based on the substitute criteria. The reasons for this are the different value added shares of the different goods or services sold, or the differing capital intensities of the different goods or services produced. For the input indicators there is a further problem: ancillary activities play no role in the determination of the principal activity. Thus, employment engaged in ancillary activities must be identified and taken out of the determination.

Output substitutes are thus the preferred basis and also easier to handle: in principle it is quite straightforward for a unit to provide data on the shares of sales for the different categories of the goods and services produced, compared to reporting on the share of employment by the different activities, for example. Having such information on the share of sales by the goods and services sold is a good basis for the derivation of the principal activity. The sales shares can be transformed into an approximation of value added shares. If there is some form of structural business statistics activity in the country, the data might be used for that transformation.

Such structural business surveys ask for data on the output of a unit, and also the cost data, which are split into intermediate consumption and components of valued added (wages and salaries, deprecation, taxes and subsidies and as a residual operating surplus). These data make it possible to calculate the shares of intermediate consumption related to sales for each of the industries. Thus, the resulting value added share by industry can be used for the transformation of the sales data into value added data. For each kind of goods and services produced the value added share of the industry typically producing that product can be applied. This is of course only an approximation as the value added structure of that unit in question could be different from the average value added structure of that industry. It is nevertheless a good approximation and can avoid the situation where goods and services sold are treated as having the same value added relations when determining the principal activity.

In addition to these general rules for the determination of the principal activity there are additional or specific rules for certain activities, such as for wholesale and retail trade or government activities, and also specific rules for various activity constellations, such as repair and maintenance, vertical integration, or outsourcing. For more information on these rules see ISIC Rev. 4.

Figure B1.1
Sections of ISIC Rev. 4

A	Agriculture, forestry and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam and air conditioning supply
E	Water supply; sewerage, waste management and remediation activities
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	Transportation and storage
I	Accommodation and food service activities
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific and technical activities
N	Administrative and support service activities
O	Public administration and defence; compulsory social security
P	Education
Q	Human health and social work activities
R	Arts, entertainment and recreation
S	Other service activities
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U	Activities of extraterritorial organizations and bodies

[72] 2008 SNA, para. 4.24.

Annex B2: Classification of institutional sectors [72]

The institutional sector classification groups together similar kinds of institutional units according to the nature of the economic activity they undertake. Corporations, NPIs, government units and households are intrinsically different from one another in their economic objectives, functions and behaviour. Where the three basic economic functions:

▸   Production of goods and services

▸   Consumption to satisfy human needs

▸   Accumulation of various forms of capital

are concerned, only the production function is of interest in the context of the SBR. Thus, only the institutional units engaged in production are covered in the SBR and can be classified by institutional sector and only households which engage in market production are covered in the SBR (in the form of sole proprietorships). As the SBR covers only resident units, the institutional sector "rest of the world" is also not of interest.

Corporations undertake production and/or accumulation, but not consumption. Corporations are split into non-financial and financial kinds. This distinction is made

because of the special role that financial corporations play in the economy. The government undertakes production, accumulation and final consumption on behalf of the population. Households undertake consumption on their own behalf and may also engage in production and accumulation. NPIs are diverse in nature: some behave like corporations and some undertake activities similar to government but independently of it.

The five institutional sectors relevant for the SBR are:

> ▸ Non-financial corporation sector (S.11)
>
> ▸ Financial corporation sector (S.12)
>
> ▸ General government sector (S.13)
>
> ▸ Household sector (S.14)
>
> ▸ NPIs serving households (S.15)

The distinction between corporation and government is fundamental, but also difficult to apply. It refers to the distinction between market and non-market production. Corporations produce for the market and aim to sell their products at economically significant prices. Government units organize and finance the provision of goods and services to individual households and to the community as a whole. They may produce most of these goods and services themselves but the products are either provided free or at prices determined by considerations other than purely market forces. Households are primarily consumers but some also produce goods and services for the market. When production takes place in the household but not in a separate legal entity owned by the household, it is described as an unincorporated enterprise. It is then part of the institutional sector households.

As mentioned above, NPIs behave mainly as non-market producers in that they produce goods and services not for the generation of income or profit. Some NPIs, however, behave like corporations and should thus be classified under S.11 or S.12. Other NPIs that are controlled by government are to be classified under S.13. The remaining NPIs, those that produce goods and services but do not sell them at economically significant prices and are not government-controlled, are classified in a separate institutional sector called "NPISHs". They are mainly non-governmental social institutions, such as churches, social clubs, charitable associations, etc.

The five institutional sectors can further be divided into subsectors, depending on the national needs and intended analysis. ESA 2010 includes subsectors for all the institutional sectors except S.15 (NPISH), based on the type of institutional unit and control criteria. The sectors S.11 (non-financial corporations) and S.12 (financial corporations) are subdivided by whether the units are public, foreign controlled or nationally private controlled. The first category would include public corporations and the second resident units belonging to foreign controlled multinational enterprise groups, whereas the last category comprises all the other institutional units classified in S.11 and S.12.

The financial sector S.12 is further divided by type of institutional units into nine subsectors, from central bank, other monetary financial institutions, to insurance companies and pension funds. The general government sector is split into four subsectors: central government, state government, local government, and social security funds. Lastly, the household sector is also split into different types of households. Only the subsectors S.141 (employer households) and S.142 (own-account workers) are covered in the SBR.

The classification of institutional sector is different from an activity classification, such as ISIC. Accordingly, the activity coding is not of much help for the coding according

to the institutional sectors. Only in certain cases does an activity code lead directly to a sector code. Examples are central banks, insurance companies, pension funds, and the activities of public administration. The legal form of an institutional unit is also a criterion that is partly of help: the main examples are the legal forms of sole proprietorship which are allocated to the household sector S.14. The legal form of non-profit associations is not, however, in itself useful as a criterion, as the NPIs can be classified to S.11, S.12, S.13 and S.15 according to their behaviour and control.

The main challenge for the implementation of the sector classification is the correct allocation of units to sectors S.12 and S.13. Sector S.14 can be implemented automatically based on the legal form and NPIs need to be coded manually. Thus, if the allocation to the sectors S.12, S.13 and S.14 is fixed, the remaining units that are not sole proprietorships can be allocated to sector S.11.

For the allocations to sectors S.12, S.13 and S.15 additional information is required that is usually not available in the SBR and in the administrative data sources used in the SBR. The complexity of these sectors lies in deciding whether a unit is a market or a non-market producer and whether the unit is government-controlled or not. Further, especially as concerns the financial sector, there are various types of units that need special attention, such as holdings, head offices, SPEs, captive financial units, and artificial subsidiaries.

Working closely with the experts of national and financial accounts in the national statistical office is recommended in order to achieve a high quality and coherent classification according to the institutional sectors. Such cooperation could also include the central bank, as often these institutions produce financial statistics and compile financial accounts.

Figure B2 provides an overview of the relationships between the different kinds of institutional units and the institutional sectors.

Figure B2.1
Institutional sector as determined by legal description and market or non-market production[73]

Type of producer		Market producers		Non-market producers	
Standard legal description		Goods and non-financial services	Financial intermediation	Public producers	Private producers
Private and public corporations[74]		S.11 Non-Financial corporations	S.12 Financial corporations		
Cooperatives[75] and partnerships recognized as independent legal entities		S.11 Non-Financial corporations	S.12 Financial corporations		
Public producers which by virtue of special legislation are recognized as independent legal entities		S.11 Non-Financial corporations	S.12 Financial corporations		
Public producers not recognized as independent legal entities	Those with the characteristics of quasi-corporations	S.11 Non-Financial corporations	S.12 Financial corporations		
	The rest			S.13 General government	
NPIs recognized as independent legal entities		S.11 Non-Financial corporations[76]	S.12 Financial corporations	S.13 General government	S.15 NPIs serving households
Unincorporated household enterprises: Partnerships not recognized as independent legal entities Sole proprietorships	Those with the characteristics of quasi-corporations	S.11 Non-Financial corporations	S.12 Financial corporations		
	The rest	S.14 households	S.14 Households		
Head offices whose preponderant type of activity of the group of corporations controlled by them is the production of:	Goods and non-financial services	S.11 Non-Financial corporations			
	Financial services		S.12 Financial corporations		

[73] Institutions in the third or social economy (TSE) sector are discussed in chapter 4, section 4.10.6.

[74] This category may include social enterprises which are part of the third or social economy (TSE) sector.

[75] This category may include the cooperatives and mutual societies which are part of the third or social economy (TSE) sector.

[76] NPIs in this category are part of the third or social economy (TSE) sector.

Annex C
Examples of statistical business registers

Annex C1: SBR development at Statistics Denmark

C1.1: Summary

The development of Danish business statistics from a situation where they covered only activities in the agricultural sector, which used to be the main activity in Denmark, to a situation where they cover all business activities, and the move from separate lists of businesses updated by individual survey statistics divisions to a comprehensive business register, illustrate how a national statistical office can become an important player in developing a business register that is useful for statistics as well as for public administration and the businesses themselves.

As long as statistics are based on censuses, the need for an SBR is not obvious. When, however, it is acknowledged that:

- The national statistical office does not have the resources to conduct censuses.
- The results of censuses are out of date before the results can be published, and
- The response burden is high.

the need for an SBR is evident. This was already recognized by the Nordic Council in 1952, but it was not until 1959 that the first version of the Danish SBR was established.

Thanks to cooperation between Statistics Denmark and other public institutions, it has been possible to influence the development of different administrative registers. The breakthrough came at the end of the 1960s with the establishment of four central administrative registers. These enabled Statistics Denmark to establish a statistical register, as well as an administrative register, based on regular electronic updating routines. A special act was passed to make a distinction between the statistical register and the administrative register. The basic information in the two registers may be the same (names and addresses, and so forth, as well as the administrative identification numbers), but some information is available only to statisticians.

In addition, technical developments have widened the possibilities for a much more flexible and comprehensive business register and for identification of inconsistencies. In 1993, Statistics Denmark established a relational database, which provided the basis for a complete separation of the SBR from the central administrative business register (ABR). The ABR, which earlier had been maintained by Statistics Denmark, was moved to the tax authorities and then to the Ministry of Business Affairs. Linking administrative units to legal units was found to be important for public purposes and the tax authorities, among others, found it useful for the ABR to include LKAUs, linked to legal units, for administrative purposes.

It is interesting that, because of its importance to the economy, the agricultural sector itself (organizations and farmers) was very interested in the agricultural statistics, making it possible to maintain an annual census with data collected locally until 1970, after which the Building and Dwelling Register of the municipalities formed the basis for the Danish Statistical Agricultural Register, separate from the SBR. The Building and Dwelling Register was used to update the Statistical Agricultural Register until 2008.

Still more administrative registers have been developed since 1970, and Statistics Denmark is, by law, required to participate in the preparatory work. Statistics Denmark determines whether the new register contains information of interest to statistics and tries to influence the legislator to take statistical needs into consideration. One recent example is monthly reporting of various wage and employment figures to the tax authorities.

Statistics Denmark has worked to remove shadow registers kept by individual statistical divisions. The latest developments have been to incorporate, first, the agricultural register into the general business register, including the use of two administrative registers covering the agricultural sector (the Central Household Animals Register and the General Agricultural Register); and, second, the use of the digital maps and transformation of information on forests from these maps to categorize LKAUs. Further details of these experiences are provided in the following sections.

C1.2: First establishment of Danish business register

From 1896 to 1958 seven censuses were carried out covering only the non-agricultural industries. The same form was used for all activities and sizes of units. This did not produce economic statistics that were fully acceptable to users. The SNA also indicated that the establishment was the relevant statistical unit to use. The municipalities distributed and collected the census forms, carried out the primary validation and sent the forms to Statistics Denmark. Even after this, Statistics Denmark still had to contact many businesses by telephone.

In Denmark the first SBR was established on the basis of the business census in 1958. Called a "statistical address list", it listed companies with activities in industry, building and construction. The list was updated manually by information from address directories made available to Statistics Denmark free of charge. The register from which samples could be drawn from an updated population was ready in 1961. At that time only punched cards were used, which made use of the list very resource-intensive.

C1.3: Breaking point one – precondition for a second SBR

By the end of the 1960s, four electronic administrative registers had been established and electronic data processing was introduced at Statistics Denmark. This was a revolution. The four registers were as follows.

> ‣ *Wage Earner and Employer Register* (1965) – this had the purpose of keeping information on the wage earners' supplementary pension.

> ‣ *Central Person Register* (1968) – every person in Denmark received a personal identification number (ID) which had to be used by the Government in all contacts with the public.

> ‣ *Value Added Tax register* (1967) – almost all businesses were obliged to register by the tax authorities in an electronic register when their turno-

ver exceeded a certain threshold. For several years, the register was used to produce statistics on turnover.

> ▸ *Income Tax Register* (1969) – it became compulsory for all employers to withhold income tax for all employees each time wages and salaries were paid. The collection of payments to the wage earners' supplementary pensions could be collected at the same time. Later, the register was extended to cover dividends, etc.

The income tax system, which was built to manage the flow of money, was not connected to the VAT system. These two new systems introduced the possibility of producing statistics on wages and numbers of employees.

One condition for the new systems to function was the establishment of a personal ID-number system. Besides the personal ID-number, the Central Personal Register also included an address register. Each occupied house and dwelling was in the address register. Each road was assigned a municipality code and a road code. The house numbers (and for flats also floor and flat numbers) that were already known were registered in the register. As businesses were not in the address register this created certain challenges. As businesses had house numbers on roads, however, Statistics Denmark decided simply to extend the official address register with these extra house numbers.

The address register is still not complete. It will be fully incorporated into the building and construction register. At the same time all businesses will be given an official address. There are some challenges concerning businesses not connected to buildings (where x-y coordinates will be used) and concerning owners who are not living in Denmark (where addresses are still needed). The current system is usable, but complete integration is not expected until in the middle of 2016.

The law on Statistics Denmark in 1966 gave Statistics Denmark the right to access to the new registers. Statistics Denmark received information from the VAT register each month, including information to calculate turnover, and it received information from the Income Tax Register, including payments to the wage earners' supplementary pensions, from which the number of employees could be calculated.

C1.4: Content of second SBR

In the first instance, Statistics Denmark did not make any attempt to integrate the two tax systems. Two separate files were kept. It became obvious, however, that it would be of interest to link information on turnover and number of employees. In addition, the number of units (initially administrative units, later legal units) depended upon which system (the VAT system or the income tax system) was the basis for the calculation, which was not satisfactory.

It was difficult to link the two registers, and a matching system had to be developed. Fortunately, both the VAT register and the income tax register kept links between two or more administrative units that were connected to the same legal unit. Each of the registers also identified one administrative unit as the main unit. This information was used to create a legal unit in each register.

If the legal unit was a single proprietor business the personal ID number was registered in both registers. Thus, in the case of a single proprietor business, the two registers' identification numbers could be connected with virtually no problem. This was also the case for partnerships. The addresses did not always correspond in the two registers, but this was not a significant problem. A decision simply had to be taken about which should be used.

For other legal forms matching on addresses was the starting point. In addition, matching on telephone number was used. As not all businesses were registered in both registers the result of matching was a group of matched units and two groups of non-matching units. During the following updating routines all units with a new or changed address were re-matched with the non-matched group from the other register, by which process the quality of the matching was improved.

C1.5: SBR versus ABR

In 1967 the Ministry of Administration asked Statistics Denmark to establish a general ABR. Statistics Denmark took responsibility for establishing a business register both for the public and for statistical purposes, but making a clear distinction between the public part and the statistical part. Identification information such as the legal form and the activity code was the same. Statistics Denmark suggested introducing a common ID number that could be used to identify each legal unit, but this turned out to be impossible at that time, especially because it would have required a change in the two main source systems (the VAT system and the income tax system). A law on the Administrative Business Register was passed by the Danish parliament in 1975.

Further development of the two Danish business registers is discussed below, but before moving to this, some attention should be given to the cooperation concerning the number of employees at LKAUs.

C1.6: Work to establish LKAUs

Statistics Denmark wanted to produce employment statistics distributed by municipality, or even smaller areas. When the income tax system was designed it was not possible to attribute the employees to a local unit, so Statistics Denmark had to conduct a survey for this purpose. As businesses already withheld individual taxes for each employee, the only missing information in administrative sources was where the employees worked. In most cases, the employer had only one local unit and all employees could be attributed to that location. This location was usually at the address registered by the tax authorities. It was also possible to see where the employees lived using administrative registers. If a group of employees lived far from the location identified by the tax registration of the business, it indicated that at least one more location was used by that employer. Thus, Statistics Denmark could limit its survey to those employers where these kinds of differences showed up.

A system to collect information on LKAUs was established together with the tax authorities and the business organizations. The system came into action in 1980. Each LKAU belonging to one legal unit with more than one LKAU was given a three-digit number. Every November, Statistics Denmark sent a list of LKAUs with addresses and activity codes to the legal unit. The legal unit had to complete the list by adding changes to addresses and activity codes by deleting LKAUs that no longer belonged to the legal unit and by adding new LKAUs. In the event of a takeover Statistics Denmark asked for information from the legal unit that had taken over the LKAU.

When a legal unit submitted information on income and withheld tax for each employee to the tax authorities, it had to indicate the three-digit number of the LKAU at which the employee was employed at the end of November. The employer could also give information about whether or not the employee had been employed for the whole year. Where an administrative unit corresponded exactly to one LKAU, the legal unit did not need to fill in the LKAU list or the three-digit number of the LKAU for the employees.

This working-place-based system also included registration of correct addresses for those legal units that had only one LKAU but at a different address from that registered in the VAT register.

C1.7: Work to establish a coherent SBR

During the 1970s and 1980s a simple record system was built. Even though LKAUs were registered for those legal units that had more than one, it turned out that there were some problems with the locations of the one-to-one units. An investigation unveiled the reason: for some units the address registered in the VAT register was not the address where the activity took place but the home of the owner or the place where the owner kept his or her account. This meant that, as the SBR was updated every month, useful information from the working-place system described above was overwritten. As a result, it was not sufficient to keep records on those LKAUs where more than one was connected to a legal unit. There was a need for a complete list of LKAUs. At the same time, as a result of technological developments, it became possible to move towards use of relational databases.

C1.8: Common identification

A committee was established by Statistics Denmark in 1970 with members from several ministries to determine how centralization of business registration could be achieved. Although its report explained the need for common identification numbers, and although this same message was repeated by a similar group in 1982, this did not move the VAT and tax authorities before these two directorates were joined. Even then the situation was not ideal as the only agreement that could be obtained was that the VAT and the tax numbers should be converted into one.

As Statistics Denmark already had done the work required to link the two systems, all that was needed was agreement about converting the identification numbers to the new number. In the case where there was a one-to-one situation, the adjusted tax number was used. In case of several administrative identification numbers being used by a legal unit, these were all converted to the new numbering scheme, but with the main tax number still identified as the main number. Now the VAT and tax numbers were the same even though, in some cases, one number was used for VAT registration and two numbers were used for tax registration. The numbers are referred to below as tax numbers.

Statistics Denmark took the new ID number into use in 1987 in the ABR. In 1992, the new number was also in use in the joined VAT and tax register.

C1.9: Third establishment of the SBR

Statistics Denmark was then able to drop its matching procedure. Instead, a relational database was created, including the administrative units and their relations and their relations to the legal unit, which was still the Statistics Denmark unit used in the ABR still being maintained by Statistics Denmark. The relational database also included the enterprise and the LKAU, and the relation between these, and the relation between the enterprise and the legal unit, and in some cases the relation between LKAU and an administrative unit.

Business demography was a new important area. Accordingly, the new SBR also included information about demographic events at LKAU level, including relations between LKAUs involved in the same event.

The process and system that followed an LKAU over time was very complicated, using information about changes in addresses and activity codes, about starting and stopping legal units, and about employees and turnover.

C1.10: Fourth establishment of the SBR

No more than six years passed before the SBR was again revised. The reason was the merging of VAT and tax by the tax authorities in 1992. In addition, the revision of the law on the ABR resulted in a wish to renew business registration more fundamentally to include other administrative registers as well and to require that the public administration should not ask for the same information more than once.

In 1999, the Central Administrative Business Register was moved to the Ministry of Business Affairs, which is responsible for its maintenance. Formally, legal units and LKAUs were transformed to the new ABR as legal units and legal LKAUs. Statistics Denmark kept legal units, enterprises and LKAUs together with the legal LKAUs and tax units in the SBR. Registration of demographic events was also moved to the ABR. However, this did not work properly in practice since it was not possible to change LKAU-identification numbers of legal units in the ABR. After a while Statistics Denmark therefore took over the demographic registration, which was then integrated in the work with the SBR.

The work of Statistics Denmark to link administrative units to legal units was crucial for the successful establishment of the Central Administrative Business Register, and the tax authorities, among others, found it useful for administrative purposes for the ABR to include administrative LKAUs.

Figure C1 illustrates the Danish SBR, the main sources and output in terms of providing sampling frames for survey-based statistics. The main sources of the SBR are the ABR and the tax register. Other statistical and administrative registers, such as that on agriculture, are also used. SBR staff members update the SBR through processing of information received from administrative sources and from surveys conducted by Statistics Denmark. Staff members from survey-based statistics divisions have access to update the SBR, on the basis of information received from their data collection.

C1.11: Further development

The SBR has been further developed by including enterprise groups. A system for registration of owners of legal units has been taken into use and the data exchange with the European Group Register has been incorporated.

One of the latest developments has been to incorporate the agricultural register into the general business register. This includes use of two administrative registers covering the agricultural sector, the Central Household Animals Register and the General Agricultural Register, and the use of digital maps to categorize LKAUs. If divisions of the national statistical office discover need for changes, these are recorded in the SBR which helps to keep the SBR updated. It also helps to ensure that statistics produced by different divisions are consistent.

A few years ago a new system for employers to report the income of their employees to the tax authorities came into use. Every month the employers have to report the income paid to each employee and by which legal LKAU the employee is employed. The information is used to produce high-quality employment statistics. It is also used to check if a legal LKAU has been taken over by a new employer in which case the corresponding statistical LKAU will be transferred to the new enterprise.

Looking ahead, it may be compulsory for businesses to report ownership of other businesses to the Ministry of Business Affairs, and this information may be available to Statistics Denmark. Statistics Denmark has asked for detailed ownership groupings (below 5 per cent, 5–10 per cent, and so on) but the actual groupings have not yet been decided.

C1.12: Lessons learned

The Danish experience suggests that three things are very important:

> ▸   Identification numbers

> ▸   Keeping things simple, having information on administrative registrations and keeping track of registration history

> ▸   Cooperation with public administrations and business organizations

Common identification

The most important thing is the identification numbering system. The optimal situation is that of one common identification number for legal units used by all organizations. This enables different administrative identification numbers and the related information to be connected through this single legal unit identification number. Groups involving relevant ministries and other organizations set up to discuss common identification of the units, including legal form and local units, have been very useful in the development of a common identification number in Denmark. Keep things simple

It is important to keep things simple. This means not mixing everything together but keeping things separate. In particular, administrative information should be kept untouched and connected to administrative units. SBR processes can then transform this information into relevant information about statistical units. It is also important to know when information has been updated, from what source, and to what point in time the information relates. If information is received about a change in activity code it is important to know if this change relates to the current year or the previous year.

Cooperation with administrative sources

Good cooperation with administrative sources is crucial to ensure progress over time. The General Director of Statistics Denmark is a member of the board of Datacentralen, a publicly-owned data processing centre which operates the administrative registers of the country. This provides Statistics Denmark with an important opportunity of influence on the development and maintenance of administrative registers which is of great value also for the SBR. The close cooperation between the SBR and the Central Administrative Business Register has also been very useful.

Updating frequency

The 1966 act on Statistics Denmark gave the organization access to various administrative registers. Statistics Denmark used to get information from the VAT register each month and from the Income Tax Register, including the payment to the wage earners' supplementary pension, each quarter. Now this information is delivered each week. Statistics Denmark also gets information about each employee connected to each legal LKAU each month, and about monthly turnover, exports, imports etc. Since 1999, when the Central Administrative Business Register was established, Statistics

Denmark has been receiving updating information when the Central Business Register is updated, although this is dependent to a large extent on self-registration.

Figure C1.1
Overview of the Danish SBR – updating and extracts
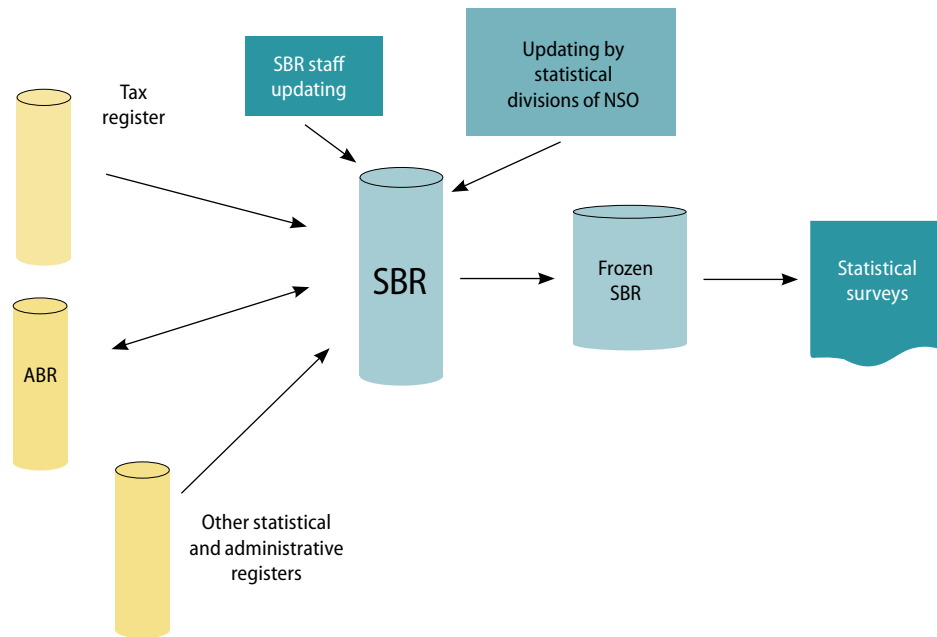
Annex C2: SBR programme of Costa Rica

C2.1: Overview

The Register of Enterprises and Establishments (REE) in Costa Rica is a structured registry of the resident institutional units (enterprises) in the private sector in Costa Rica, and their local units (establishments), engaged in the production of goods and services. It includes information on the characteristics of each unit, such as identification (ID) characteristics, location, economic activity and size.

Institutional units are enterprises associated with one or more legal units. They have unique ID numbers assigned by the National Institute of Statistics and Census (INEC). Each enterprise is further divided into one or more local units (establishments). Each legal unit of an enterprise can be further associated with one or more of the enterprise's local units. On the other hand, every local unit of an enterprise is associated with at least one legal unit. Thus, two or more local units of the same enterprise can be associated with the same legal unit. This is illustrated in figure C2 below.

C2.2: Updating process

The process for updating the REE relies on three mechanisms. First, telephone interviews are carried out to obtain basic characteristics such as identification characteristics, location, size, and economic activity. This mechanism is supported by a customized information system developed in-house that allows updating of the REE database during the interview itself.

The second updating mechanism consists of the use of administrative records from both internal and external databases, including:

    ▸   *External sources*: Ministry of Finance, Agency for the Promotion of Foreign Trade, Electoral Register, National Registry and Social Security Register.

    ▸   *Internal sources*: Consumer Price Index and Construction Price Index databases, and also the National Enterprise Survey.

A third updating mechanism, implemented since 2012, is the collection of data based on an enumeration exercise that covers one district at a time. This enumeration of local businesses is supported by the use of hand-held personal digital assistant*s* (PDAs) – computer devices – to update all the key characteristics in the REE. The main purpose of this exercise is to assess the coverage of the REE and the increase or decrease in the number of businesses in the district.

In addition to the three methods described above, special data-collection projects that are carried out by INEC for other government agencies provide inputs to update the records of the REE.

Approximately 40 per cent of the records in the REE are updated every year. Priority is given to those enterprises that have not been interviewed for a long period of time. This process of continuously updating the REE is crucial in providing an accurate picture of the current economic situation of the country and an adequate sampling framework for statistical surveys. Having a central business register avoids duplication of work and the proliferation of fragmented data sets collected by different users or agencies on an ad hoc basis.

C2.3: Unique identifier

Costa Rica has a unique identifier for each natural person and for each legal entity. The ID number for natural persons ("*cédula física*") has nine digits and is required for all official administrative procedures related to social security, obtaining a passport, and other requirements.

The ID number for legal persons ("*cédula jurídica*") is assigned by the National Registry to legal entities for paying taxes, complying with social security regulations, engaging in buying and selling operations, and other procedures. It has ten digits; the first four digits identify whether it refers to a corporation, an association, a cooperative, a foundation, a foreign enterprise, etc.

C2.4: Administrative records

Social Security Register

The social security register (SSR) consists of each month's payroll, with details on the number of employees, salaries, and economic activity. It includes information obtained from employers, self-employed persons, individual employees, and persons employed under a special (collective) contract.

The SSR is used to improve the coverage of the REE by means of a data integration process. The first step is to reclassify data from the SSR database of employers that refer to geographical areas and economic activities (according to ISIC Rev. 3) to the INEC statistical classification of territories and ISIC Rev. 4. Once this reclassification is done, an assessment of the SSR database is carried out to determine which characteristics are relevant for the REE.

Structure of business register
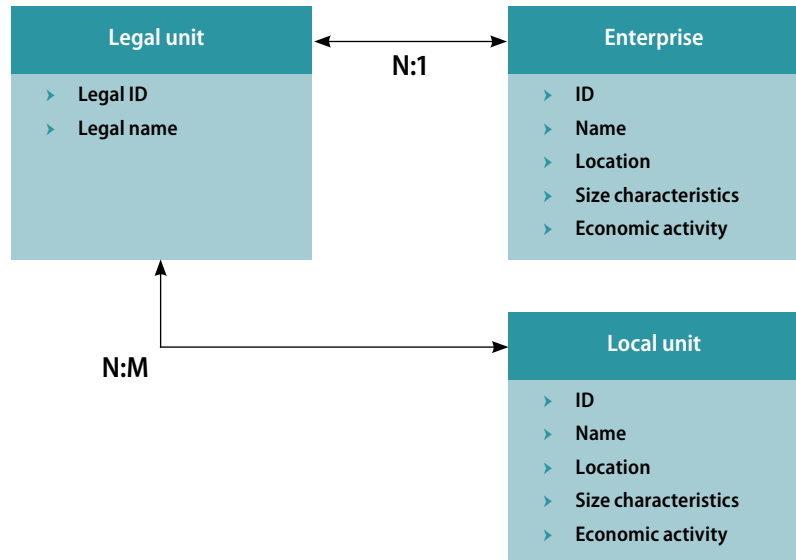
Lastly, the SSR database is linked with the REE database using either the unique ID numbers (for natural or legal persons) or the passport number in the case of foreign persons. Non-matching records are analysed in detail to check whether the enterprise is already included in the REE under another ID number (to avoid duplicates), or whether it is inactive. Those records from the SSR database that are not found in the REE database are added as new records, and are then immediately updated to the extent possible by means of follow-up phone interviews.

Records from the Ministry of Finance

Administrative records from the Ministry of Finance contain data on the taxable income of natural and legal persons, over a specific period of time, collected by the Tax Revenue Administration. This information is used for two specific purposes: first, to update the data of large taxpayers and improve the coverage of the REE by adding new records for large enterprises whose ID numbers are not found in the REE; and, second, to identify large or complex enterprises based on their level of taxable income and their number of employees by economic activity.

Register of the Agency for the Promotion of Foreign Trade

The Trade Intelligence Division of the Agency for the Promotion of Foreign Trade (PROCOMER) provides INEC with its register of exporting enterprises every year, within the framework of a joint programme to carry out a Census of Exporting Enterprises. The goal of this programme is to measure the number of jobs created by the exporting sector and to assess the share in total exports of different categories of enterprises by size.

The PROCOMER register of exporting enterprises is used to improve the coverage of the REE on the basis of a comparative analysis that allows identification and addition of new records for exporting enterprises not found in the REE.

National Registry and Electoral Supreme Court databases

The National Registry is the institution responsible, among other things, for the cadastre and registry of real estate, industrial, and other property, and also for the official registration of legal persons. The Electoral Supreme Court is in charge of civil registration, namely, registration of all events that are relevant from the point of view of civil law (births, marriages, divorces, deaths, emission of personal ID number, and so forth). These two databases are used for cross-validation purposes. Specifically, they are used to verify the existence of the unique ID numbers (for natural and legal persons) of the enterprises included in administrative records, and to verify whether the legal name of an enterprise in the REE corresponds to the name assigned by the National Registry or the Electoral Supreme Court.

Consumer Price Index and Construction Price Index databases of INEC

The list of enterprises providing data to INEC for the Consumer Price Index and Construction Price Index programmes is periodically used to update the REE. With the help of an information system, the data collected through the price surveys are compared to the information in the REE database, thus avoiding the need to conduct additional phone interviews.

INEC National Enterprise Survey framework

The National Enterprise Survey is conducted on a quarterly basis on a sample of enterprises from the private sector engaged in various economic activities throughout the country. It collects data on the number of enterprises, jobs, hours worked, and salaries. Its main purpose is to provide information on the situation of the Costa Rican labour market from the point of view of the enterprises.

It also provides an opportunity to update the information of enterprises belonging to the same enterprise group. In some cases, it is possible to obtain from a single interview data on the various individual enterprises that belong to the same group, even if not all of them are included in the sample. This also helps to improve the coverage of the REE by identifying the enterprises that are not yet included in its database.

The REE is updated with data collected through the National Enterprise Survey with the help of a special module, which allows comparison of basic data of the surveyed enterprises with the current contents of the REE database.

C2.5: Cooperation agreements with sources of administrative records

An inter-agency cooperation agreement between INEC and the Ministry of Science and Technology provides for the electronic exchange of data and mutual technical assistance in respect of the collection, use, and maintenance of statistical data for the national survey for the compilation of indicators on science, technology and innovation.

Within the framework of an agreement between INEC and the Institute for Research in Economics (IICE) at the University of Costa Rica, INEC provides the sampling framework for the survey on business expectations. The sample includes businesses

from agriculture, manufacturing, construction, wholesale and retail trade, and other service activities.

A number of other inter-agency agreements have been established between INEC and the Ministry of Finance, the Agency for the Promotion of Foreign Trade, and the Ministry of the Economy, Industry and Trade. These agreements, which reflect the good relationships that exist with these institutions, allow exchange of information to improve the operation of the REE programme.

Annex C3: SBR of Statistics Canada

C3.1: Introduction

When Statistics Canada redesigned its SBR system over the period 2007–2008, it built a Microsoft Windows application installed on a client PC using a service-oriented architecture. There are five major components making up the system, namely:

- ▸ VB.Net is the programming language for the Windows Forms user interface, the business layer and the data layer.
- ▸ SQL Server 2005 is the underlying database that both stores and manipulates the data.
- ▸ SAS is used to crunch and manipulate input data from external sources.
- ▸ The system is message based and uses BIZTALK for routing the messages.
- ▸ Web services are used to manage both security and access to the data.

All users access the SBR by means of the same common interface with a privilege administration tool as the control mechanism to manage this access. The SBR includes information on legal and operating units and their structural relationships. It maintains a journal and log of all updates applied to these units and to stratification characteristics and information on reporting arrangements.

Eight modules within the register manage different aspects of this information.

C3.2: Browser module

The browser module enables the user to browse and search for information on a given enterprise. It displays information such as the business structure, collection entities, the response burden, and the history of updates contained in the register (journal, log and snapshots).

The log contains all the updates performed on a given variable. The journal records significant events, such as amalgamation and dissolution, concerning a given business. The monthly snapshots show the image of the business structures at regular points in time in the past (previous months).

Figure C3.1
SBR environment
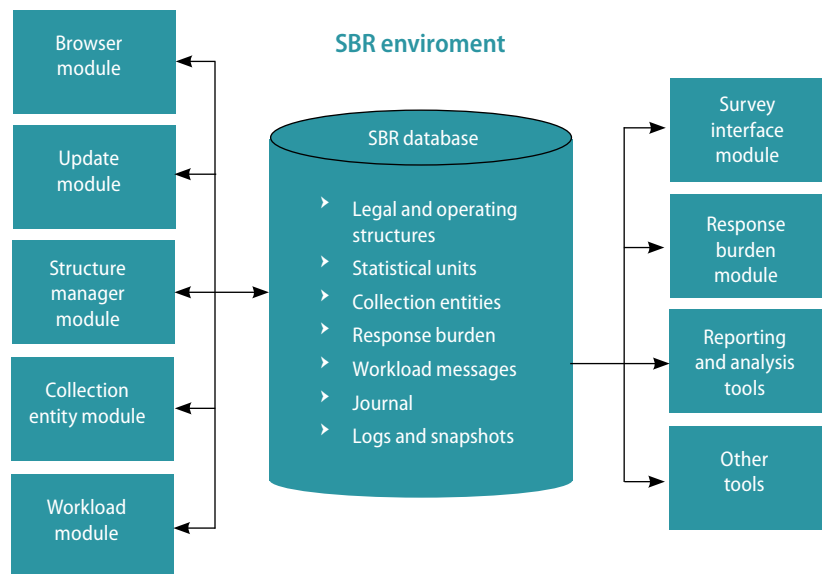
C3.3: Update module

This component serves to control and manage all updates, both manual and batch, that need to be applied to the database. The SBR receives requests or signals for updates from various sources such as subject matter, collection and external administrative sources. Although each may follow a slightly different process, in general a request is vetted to determine if human intervention is required (workload) and if accepted, it is applied to the database. Whether a request is rejected or accepted, the systems will send out notification to ensure that the affected parties are aware of the status of the request.

C3.4: Structure manager module

The structure manager module is used to show complex structures via the browser. It manages these enterprise structures and the links that exist between production entities. It manages and controls the parent–child links, propagates attributes within the structure, and checks the coherence of the structure once changes have been made.

C3.5: Collection entity module

This module is responsible for generating and updating collection entities (containing information used in contacting respondents) based on the information contained in a survey control file received from the survey's sampling process. It also manages the manual customization of collection entities that is performed by survey managers based on pre-established business rules.

C3.6: Workload module

When an update signal arrives at the SBR that requires manual investigation or verification, the update module generates a signal to the workload module indicating the need for review by an analyst or profiler. The workload module manages, prioritizes

and assigns the signals to the analysts or profilers. After manual review of a signal, the analyst or profiler either implements or cancels the corresponding change request.

C3.7: Survey interface module

The function of this module is to produce two key SBR products that are necessary to SBR partners in conducting their surveys: the frozen frames for sampling; and respondent information file for collection.

The monthly standardized frozen frame is the result of extracting all units that comprise the total business population. It contains a list of all units of production with their tombstone information, the industrial classification, the detailed geographical code, the size variables (such as revenue and employees), and other information to satisfy sampling procedures. Survey methodologists use this file primarily for the generation of survey samples. It is also used by subject matter divisions as an input to their edit, imputation and estimation system.

From the units selected for a given survey by survey methodologists, the SBR provides the respondent information file to collection staff. This file contains the information required in order to carry out data collection, such as the contact name, address and telephone number.

C3.8: Response burden module

This module presents all information relating to respondent burden for economic surveys. It displays information about all contacts that the national statistical office has had with any given enterprise. The response burden tool displays this information by survey, enterprise, contact name, and questionnaire. It also calculates the actual burden per enterprise. Lastly, it provides extractions concerning exclusion orders and cases that require specialized treatment. A central frame used by the entire economic survey programme means a truly comprehensive view of response burden and thus facilitates its management.

C3.9: Reporting and analysis tools and other tools

These tools produce the reports needed to manage survey operations and analyse subpopulations. The survey frame assessment tool extracts and presents changes that have occurred in a survey's population between two reference points. It includes births, deaths and changes to industry codes and size indicators such as revenue. The survey frame assessment tool can dynamically display all changes that occurred as of the previous day for selected characteristics by operating entity. This is of particular importance as changes can be reviewed immediately prior to the production of the monthly frozen frame. Other tools include the analysis of updates to the SBR and demographic analysis of the business population.

Annex C4: Business register at the National Statistics Office of Georgia

Activities for the development of the Business Register at the National Statistics Office of Georgia (GEOSTAT) started with evaluation and identification of gaps and specific problems. The evaluation resulted in the identification of several serious gaps, such as an incomplete database, several missing characteristics, outdated software (Paradox database developed in 1990s, with no possibility of adding new fields), lack of database

structure, and inefficient updating procedures (receiving data on paper and manual data entry).

At a later stage it was necessary to identify possible internal and external sources for updating the SBR and to obtain full access to the relevant administrative sources. In this regard a number of working meetings were held with various public institutions that provide administrative data. In addition, in response to the GEOSTAT initiative, the Government adopted a resolution, according to which the administrative authorities are obliged to provide regularly required information to GEOSTAT. According to the resolution, approved in February 2011 and amended in July 2011, the Revenue Service is obliged to submit monthly data to GEOSTAT on active taxpayers that are enterprises. In this context, an enterprise is considered active if it has indicated:

- Turnover in any kind of declaration (VAT, revenue, etc.)
- Number of persons employed in any kind of declaration (VAT, revenue etc.)
- Wages or number of persons employed in a wage notification
- Profit or loss in a profit declaration
- Any kind of tax payment with the exception of property tax

Active enterprises report data on turnover and number of employees to the Tax Office on a voluntary basis. Thus, the Tax Office does not receive these data from all enterprises. It submits the data it does receive to GEOSTAT.

As a result of these activities, GEOSTAT has direct access to administrative sources. In most cases, high-quality statistical products are obtained by combining administrative and survey data.

Figure C4.1
External and internal sources
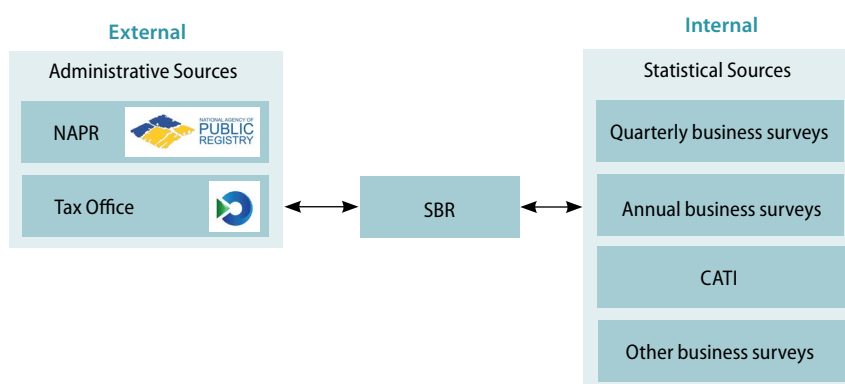
Other sources for updating the SBR include various business statistics surveys. Quarterly and annual business surveys provide information on contact details and type of economic activities of the surveyed entities.

International experts were engaged in the above-mentioned activities in order to ensure that international recommendations were taken into account and relevant methodologies for the formation of SBR in line with international standards were introduced. In this respect, cooperation with Statistics Sweden is noteworthy and greatly contributed to the development of the SBR.

The Cooperation Agreement between GEOSTAT and Statistics Sweden was signed in June 2011. The project was funded by the Swedish International Development Cooperation Agency and covered a seven-year period. One of the components and goals of this cooperation was improvement and development of the SBR in Georgia. As a result of this cooperation with the Swedish experts, an SBR maintenance strategy has been developed that covers use of administrative sources, use of survey results, assessing needs of users, planning of updates, and rules for dealing with SBR information.

Various administrative and internal data have been collected, harmonized and processed, resulting in the formation of a new database, namely the SBR. At the moment the database contains approximately 723,000 entities (both active and non-active entities). It includes all registered economic entities, including physical persons. There is almost full coverage of legal entities, but full coverage of local units, LKAUs and enterprises has not yet been achieved.

Based on information received from the Revenue Service, GEOSTAT has developed a methodology for creating a sampling frame for business surveys. This sampling frame covers active enterprises from the non-financial corporation sector. It may be used for all business surveys in Georgia.

The structure of the SBR, as shown in figure C4.2, is in accordance with international standards and recommendations.

Figure C4.2
Structure of the business register
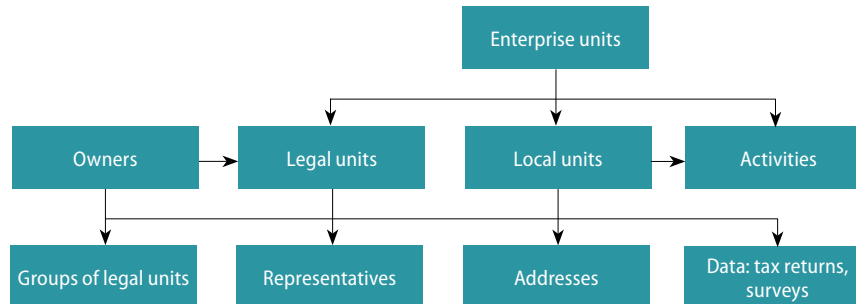
One of the major factors in the development of the SBR was the software and, to this end, new user-friendly software has been developed. The database is operating in MS SQL Server and interface in Web application (PHP).

It was also important to establish a flexible and automated system for updating the SBR. The main sources are administrative sources, namely the National Agency of Public Registry (NAPR) and the Revenue Service.

> ▸ From the Public Registry, GEOSTAT receives information on newly registered and liquidated (closed) companies, as well as information on changes.

> ▸ From the Revenue Service, GEOSTAT receives information on active taxpayers.

Update procedures are automated. The information is received from NAPR and the Revenue Service monthly, using automatic interfaces that the GEOSTAT has with the databases of these institutions. From NAPR GEOSTAT receives information about newly registered and liquidated companies, also information on changes. From the Revenue

Service, GEOSTAT receives monthly information on active taxpayers. Received data are cleaned and structured. The statuses of the companies are identified. This information received from the administrative sources is automatically reflected in the SBR.

GEOSTAT also uses results of various business surveys to update the SBR. It conducts special surveys that are directly aimed at updating the SBR and improving its quality.

Figure C4.3 shows the current SBR update procedure, and figure C4.4 shows the data flow.

Figure C4.3
Update procedures

Using administrative sources from National Agency of Public Registry and Revenue Service (monthly)

Using quarterly and annual business statistics survey results

Based on computer-assisted telephone interview (CATI)

Business register is updated:

BR

Update procedures are automated
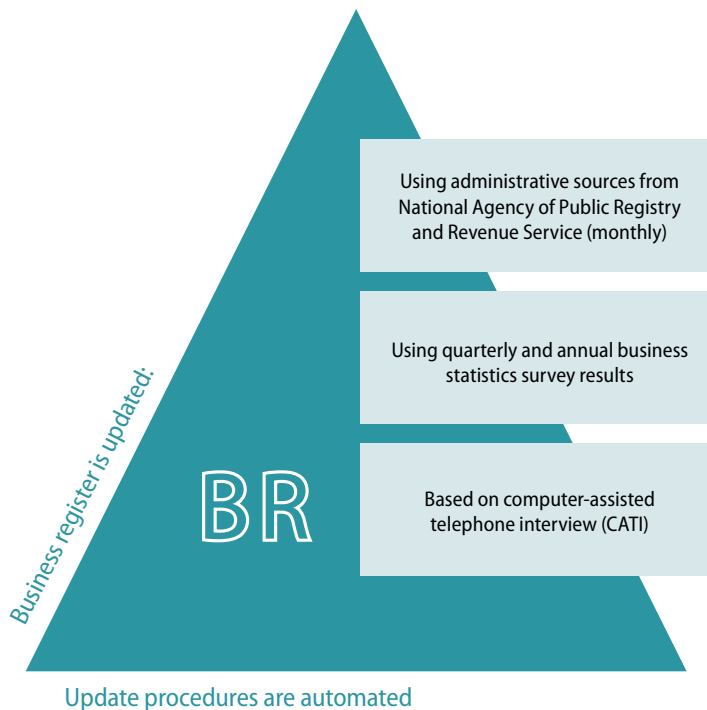
Figure C4.4
Data flow of the business register

Data source

Public register
Revenue service
Customs
Surveys
CATI
Investigations

Data processing

Automated update procedures

Databases

Public register
Business activities
Classifiers
Physical persons register (BR)
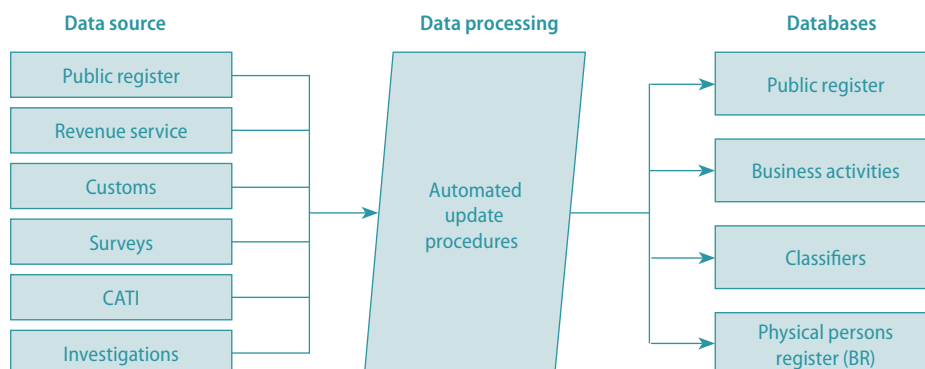
In the data received from administrative sources the actual address and type of economic activity are specified only for a limited number of enterprises. Sometimes the indicated information on economic activity is incorrect. This was a major problem for the SBR and hence for all business surveys. As a result, GEOSTAT conducts monthly computer-assisted telephone interviews for the companies in the SBR to update information about the status, kind of economic activity, and actual address of each local unit.

Relevant documentation is also essential for the users of the SBR and is a part of the overall quality of the service the SBR provides. Thus, GEOSTAT created a metadata document for the SBR that contains comprehensive information about the SBR and brief description of the characteristics, production, updating and maintenance procedures, etc. In addition, it contains information about the rules for dealing with SBR.

The relevant division in GEOSTAT is continually working to maintain the SBR and to improve its quality. Figure C4.5 shows the stages of SBR development.

It is of major importance to share experiences with other countries and improve capacity of staff through participation in trainings and workshops.

Figure C4.5
Stages of business register development in Georgia
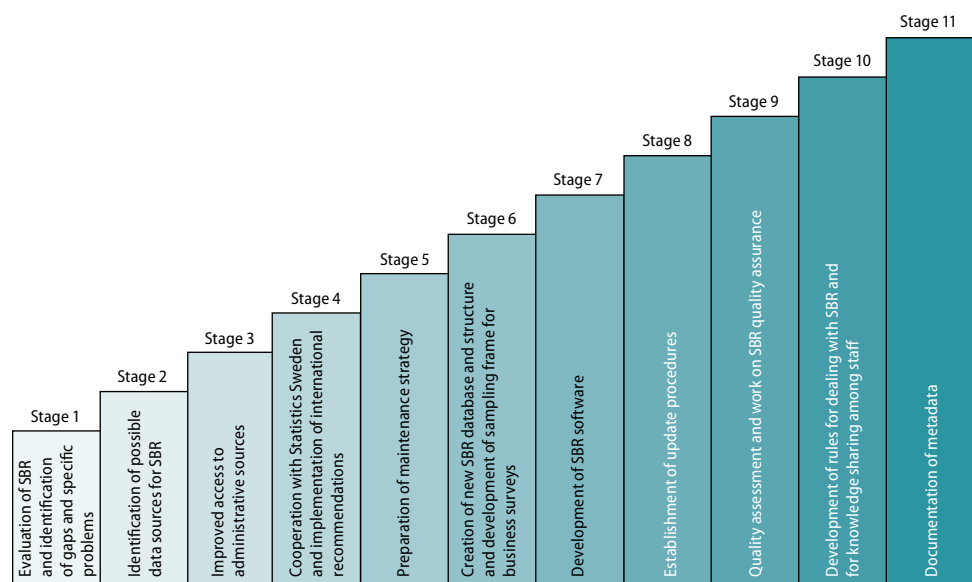
Annex C5: Redeveloping the SBR of Malaysia

C5.1: Brief history of the business register in Malaysia

The Central Register System (CRS) of Department of Statistics Malaysia (DOSM) was set up in 1994 in a PC-based system using Dbase IV. It was upgraded in 1996 into a local area network (LAN) environment. Further enhancements were made in 2002 with an upgrade to a wide area network (WAN) environment, together with the incorporation of features needed by users, requirements for new surveys and accommodation of new industrial codes.

C5.2: Streamlining business processes

DOSM has implemented an integrated statistical systems framework (ISSF) to ensure the quality in the statistical production process. This is an integrated online system to streamline statistical business processes and replace non-integrated systems. It is in line with the GSBPM developed by ECE, which provides a basis for standard terminology on statistical metadata systems and processes (fig. C5.1).

The ISSF is a flexible tool for describing and defining the set of business processes needed to produce official statistics. It can also be used as the basis for quality assessment of censuses, surveys, administrative data and data from other sources. The processes have been articulated within the ISSF module, which can be accessed by the head office and by State offices.

The ISSF provides the main business functionalities of DOSM for pre-collection, collection, processing, analysis and dissemination. The web browser will be the front-end interface to this integrated system and will provide wide access capabilities anywhere and at any time through web-based application. The business register which is known as the establishment/enterprise (EE) frame will support the pre-collection and collection module. The modules for implementing the business register are shown in figure C5.2.

Figure C5.1
Integrated statistical systems framework
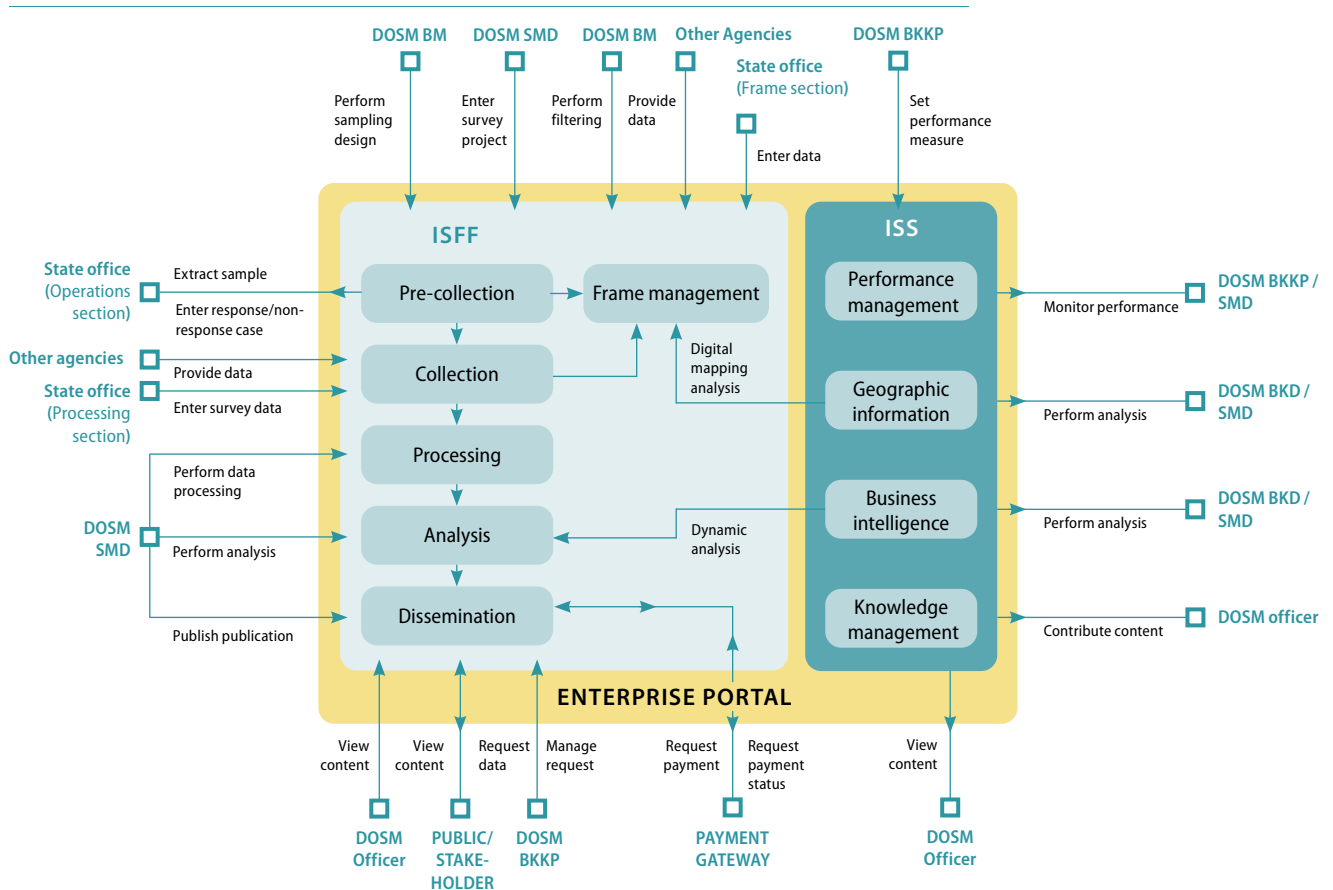
Figure C5.2
SBR modules and functions

Module	Function
Screening Agency information Business profiles	Maintenance through list information from administrative sources
Establishment/enterprise Duplicate Data collection Respondent management control	Maintenance of establishment and enterprise information
Report Business Intelligence (BI)	To generate monitoring reports and analysis tools
Geographic Information System (GIS) – EE	To monitor the number/type of establishments in an enumeration block

C5.3: Integration of the SBR with the collection system

In the ISSF platform the SBR is integrated with collection processes through the operational control information (MKO) module. The MKO module is being designed to enable subject matter experts and enumerators to update basic business information in the frame before clean data from the survey or census are processed. This increases the efficiency and timeliness of the updates. Comprehensive, reliable and timely SBR coverage and content increase data-collection efficiency and assist the State offices in performing their updating activities.

The first benefit is the accessibility to enumerators of the most recent and detailed information in the SBR. Having access to comprehensive and timely information facilitates interpretation of information received from the respondents. It assists the department in improving respondent relations and improved respondent relations lead to more comprehensive and better quality responses.

C5.4: Future challenges

Enhancing the use of administrative data sources

While data from primary sources continue to play a leading role in producing economic statistics, the use of administrative data is increasing. Thus, efforts must be made to enhance the cooperative arrangements between DOSM and administrative agencies. Furthermore, the DOSM must address barriers in terms of quality associated with the use of administrative data as basis for official statistics with respect to coverage, timeliness, frequency, validity, reliability, consistency, legality, and confidentiality constraints.

The DOSM should seek new sources of administrative data. In this respect, one of the steps to be taken is to engage with agencies using the Malaysia Corporate Identity Number (MyCoID) that was introduced by the Company Commission Malaysia. MyCoID refers to the company incorporation number which is used as a single source of reference for registration and transaction purposes with other relevant Government agencies. These will be the future platform for coordinating and acquiring data from these agencies.

There is also potential to replace some survey data with administrative data obtained from other sources. These would reduce the burden on respondents and the operational costs incurred by the office. Awareness among administrative agencies is vital in developing the uses of their data sources. This can be done through meetings, semi-

nars, training events and conferences. DOSM is monitoring the quality of incoming administrative data through ISSF.

Keeping the SBR accurate and up to date

The SBR should provide information that is as close as possible to the situation in the real economic world. Ways in which to do this include:

- ▸ Improving access and use of identification and matching data
- ▸ Improving updating from survey feedback
- ▸ Enhancing the use of administrative data sources
- ▸ Establishing area frame development and maintenance procedures

Reducing respondent burden

The SBR contains many records as it is the main medium for survey frame creation and samples selection. It should enable control of the number of establishments and enterprises that are involved in censuses and surveys. Integration between the SBR and the MKO will assist the enumerators to better manage the respondents and ultimately to ensure that the statistics collected are timely and reliable.

Business demography

There is a growing demand for business demographics, which involves statistics for specific events such as birth, death, survival, and growth. Together with the Company Commission Malaysia, DOSM is conducting research on the relevant terminology, scope and coverage and data availability.

Business profiling

Business profiling is the method for analysing the legal, operational and accounting structure of large enterprises and enterprise groups. It provides better understanding of the complex structures of the enterprise groups in the country. Profiling can be viewed as part of a broader, coordinated strategy for improving agency's economic statistics. DOSM has taken steps in this direction by creating an enterprise module in the ISSF system that can facilitate profiling activity.

Annex C6: Development of the SBR in Indonesia

First five years of SBR development in Indonesia

Badan Pusat Statistik (BPS – or Statistics Indonesia) started the development of the SBR in 2013 with technical assistance from the ABS, continuing until 2018. The general achievements of the SBR are the development of the SBR database based on the business directories for subject matter areas and the results of the 2016 economic census, the development of the SBR application system, and the use of SBR data as the source for the 2016 economic census prefilled questionnaires. The details of the development can be described as follows.

Activities in 2013

In 2013, ABS started to provide SBR technical assistance to BPS. The Australian delegations gave the initial technical assistance to BPS for the development of the business register.

At the same time, the development of the business register in Indonesia, initially called the Integrated Business Register (IBR), was carried out through cooperation between the Directorate of Census and Survey Methodology Development, the Directorate of Expenditure Accounts, the Directorate of Statistical Information Systems, and the subject matter areas (SMAs).

The Directorate of Census and Survey Methodology Development and the Directorate of Expenditure Account mainly worked as the SBR development team, creating a road map and action plan, coordinating the data gathering from the related SMAs, and doing the statistical unit concept study for corporations in Indonesia. The SBR development team members came from across BPS and divided into subteams: a steering team, a responsible persons team, a secretariat team and a sectoral implementation team. This team structure was signed-off by the head of BPS.

The study of the statistical unit concept was a big project. It involved a literature study (primarily from the 2008 SNA), exploration study and profiling, administrative data study, and field check. The literature study gave insights into the possible statistical unit coverage in SBR, consisting of enterprise group, enterprise and establishment (see fig. C6.1). After that, profiling was carried out to gather the characteristics of the real businesses in Indonesia and the structure of the statistical units. To determine the scope of the study, Globe Asia Magazine was used. The magazine provided a list of the top 100 largest private businesses or groups ranked by annual revenues that were very useful for that purpose.

A number of lessons were learned from the profiling exercise. Each industry (manufacturing, financial services, mining, real estate, and so forth) has different statistical unit characteristics. For example, the telecommunications industry has galleries; the real estate industry has toll roads, hotels, housing, office buildings; the banking industry has branch offices, sub-branch offices, cash offices, sharia offices; and so on. For that reason, it was necessary to define how the concept of enterprise, establishment, and supporting units can be applied in each industry.

Figure C6.1
Coverage of statistical units and institutional units in the SBR
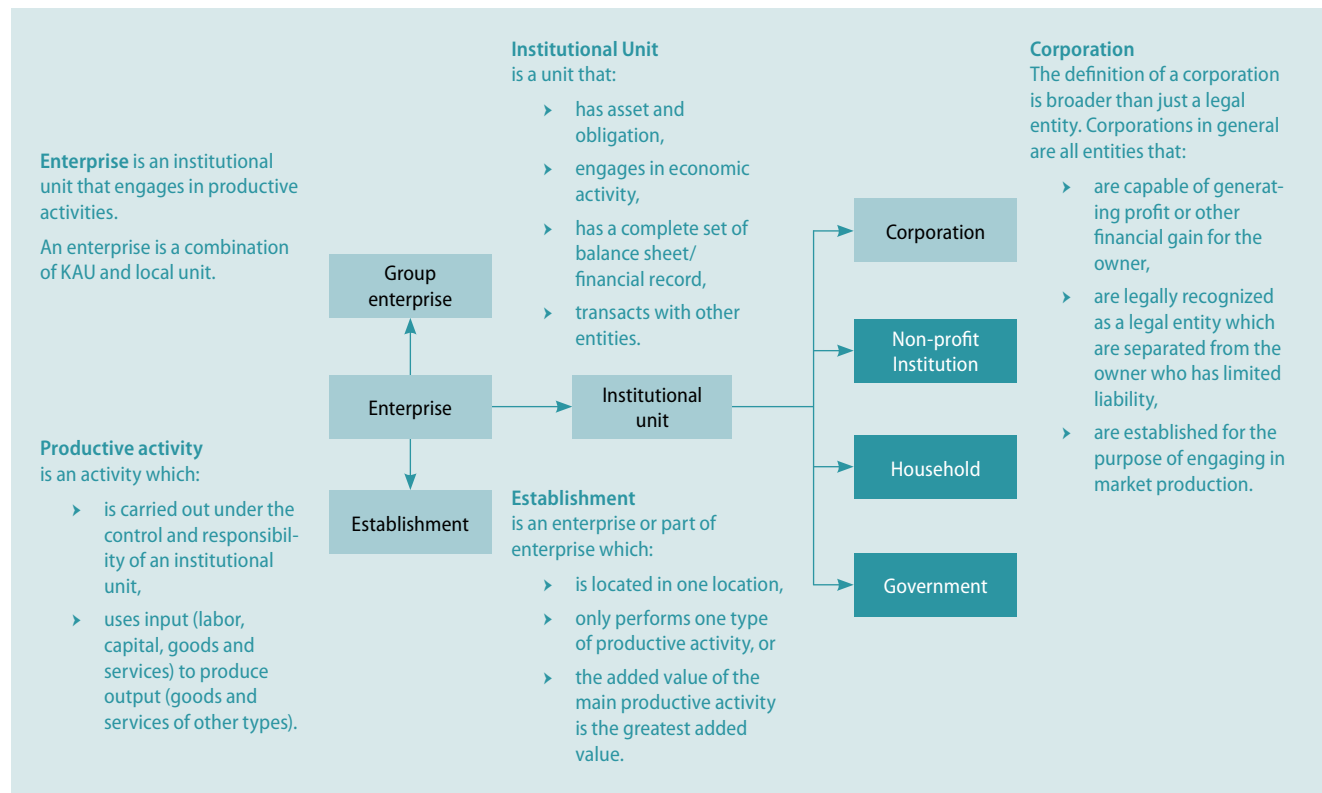
Note: Dark green shading means that the units have not yet been implemented.

Several conclusions were drawn from the study of the statistical unit concept:

> The structure of an enterprise group can be very complex so that it can have up to seven levels of subsidiaries. An enterprise's control of its subsidiaries is based on the percentage of shares it holds. Rapid stock movements need periodic updates to the get data according to current conditions.

> The address information of the subsidiaries is difficult to obtain from profiling results except for the subsidiaries that are incorporated as public limited liability businesses.

When profiling companies the following has to be taken into account:

> Size of the economic units: Given the large number of economic units in Indonesia, BPS needs to set limits on the size of the businesses to be profiled, so not to lose sight of the economic units that have a large contribution to domestic output and can reflect the real economic situation.

> Complexity of the economic units: In creating large and complex economic unit profiles, certain strategies can be used to minimize undercoverage and overcoverage, reducing the burden of economic stakeholders in censuses and economic surveys, and streamling the reporting process.

> ‣ Relationship management: In the implementation of profiling, BPS needs to build a good relationship with the businesses so as to improve data quality and timeliness of data collection.

> ‣ Profiling frequency: In the profiling activities, attention needs to be given to the possibility of changes to the real-world structure and use the intelligence business to keep the profiling results up to date.

> ‣ Special cases: Exceptions should be treated on a case-by-case basis.

Based on the description of the reports from various visits to 10 ministries as the possible data sources for the IBR, it was found that each ministry has administrative business data in accordance with the needs of each agency. Generally, the information on that was obtained from the study were:

> ‣ The type of business data that were maintained as well as the governing laws and regulations

> ‣ The characteristics and variables covered

> ‣ The authority and confidentiality of the administrative data and the opportunity for the cooperation of data exchange with BPS

> ‣ The process of maintenance of the data

A field check was also done in 2013 with a target sample of 48 businesses. From this field check, several lessons were also learned and are described below:

> ‣ Data gathering should be preceded by an official letter and a list of questions to the management of the business. The head of the business will then proceed and make the appointment. Survey officers make appointments and can only conduct interviews and retrieve data from the designated person.

> ‣ In gathering information from the business, it is important to have some initial information about the unit's status, for example, whether it is an enterprise or establishment. Based on this preliminary information, different questionnaires can be designed according to the characteristics of information that can be obtained at each of these businesses with the aim of minimizing unanswerable questions.

> ‣ It is important to provide the business with information on the purpose, the benefits (especially for the business), and the confidentiality of the survey results so that the business understands the importance of the data they provide. It was also recommended that BPS provide feedback, which may be publications or information on survey results that are accessible, especially those related to the business's contribution to the production of such data. Those are important to build good relationships and trust between BPS and businesses.

> ‣ The identification of enterprises and establishments across industries varies considerably and is quite difficult for field staff if they are not provided with real and easy guidelines for their work in the field. For that purpose, it is useful to develop concepts and enumeration guidelines for officers which are accompanied by examples.

> ‣ Businesses, especially enterprises, are more open to receiving surveys or studies conducted by the "leaving the questionnaire" method, where the questionnaires are completed by the business rather than with the interviewee. Thus, it is necessary to design a clear questionnaire. In addition,

> e-survey methods should also be considered to allow direct data to be received by BPS without the process of sending hard copy.

> ➤ The implementation of the survey or study activity takes a longer time. Based on the results of the study where the questionnaire was left at the business, at least one week after the questionnaire was received by the business, the new businesses could be contacted or, otherwise, the business contacted the BPS to return the completed questionnaire.

> ➤ Interviewing activities require at least four visits, namely: sending a notification by an official letter to the head of the business; meet one person from the business and get a contact person and make interviews; visits for the interview (two visits).

At the same time, since the end of 2013, the Directorate of Statistical Information System has started building the design and implementation of SBR database and system. The development of the SBR database was prepared in three steps, as detailed below.

Step 1: A literature study of the variables. This was done with the purpose of studying business register variables from international recommendations, among of them the Business Registers Recommendation Manual (Eurostat, 2010) and the technical assistance material from the ABS, Business Register Unit.

Step 2: A preliminary discussion with the SMAs. The results of the discussion were the following:

> ➤ Variables and characteristics for the SBR development may differ based on the statistical unit: establishment, enterprise, and enterprise group.

> ➤ The content of the common variables, the variables that were commonly shared among SMAs, and the, so to speak, "nice to have" variables were not complete in the SMA data, so that it needed to be updated for the sake of the quality of the SBR data.

> ➤ The codes and classifications of the common variables still varied in respect of the terms, units or contents across SMAs. For example, the statuses of the businesses in the agricultural sectors consisted of "1" for parent businesses, "2" for businesses without branches, and "3" for businesses with branches. Meanwhile, for manufacture businesses, the parent businesses were coded as "1", and "2" for the plants. That discrepancy also applied to the legal status codes. This necessitated the introduction of standardization for the codes and classifications to be used across all SMAs.

> ➤ There was no single ID to refer to the same unit across the SMAs. Each unit had a specific number that was not yet integrated. A specific ID was required for the purposes of data integration.

Step 3: A data review and comparison of data variables from many sources, such as the 2006 economic census data, the directory of business managed by SMAs, and the agricultural directory of the 2013 agricultural census.

Next, the design and the implementation of the SBR system were carried out to facilitate three primary activities in developing the Integrated Businesses Register. These were: matching activity (to match businesses that were maintained by the SMAs with the businesses in the 2006 economic census list as the single source); profiling; and updating activity and accessing online businesses register by related parties.

As for the SMAs, in 2013 the obligation was to share their business data with the IBR Secretariat.

Activities in 2014

Profiling in 2014 was limited to the 100 largest business groups in Indonesia based on their revenue in 2013, as reported by Globe Asia magazine for private businesses. Where State-owned enterprises are concerned, the preparation of business profiles refers to the directories of State-owned enterprises from the Ministry of State-owned Enterprises. This activity was performed by new staff members from the State College of Statistics.

In 2014, the statistical unit concept study was also continued. This study aimed at implementing the concept of the statistical unit by field-checking the profiling exploration results to identify best practices in the corporate institutional sector. This was conducted in the special capital region of Jakarta (DKI Jakarta), covering the 100 biggest private enterprise groups and State-owned enterprises in Indonesia, according to Globe Asia magazine 2013. The data collected include the general variables and the essential variables and the relationship between the parent and the subsidiaries.

During 2014, the harmonization of the industry coverage for SMA was also carried out. The aim of this work was to enable each unit in the SBR to be mapped to one SMA that would be in charge of maintaining the data. In 2014 too, for the first time, the SBR was used as the frame for the business tendency survey.

Activities in 2015

For continuous maintenance and updating of SBR data, the profiling activity in 2015 involved SMAs as the profilers, where the SMAs formed a profiling team for each work unit.

Besides profiling, there were also activities relating to the preparation of the 2016 economic census. BPS provincial offices gathered data from local administrative sources for integration into the SBR database. BPS provincial offices also helped in completing the administrative area information of the SBR units so that they could be used for the prefilled economic census questionnaires.

Activities in 2016

In 2016, there was a change in the name of the business register from integrated business register to (SBR). The activities in 2016 are grouped into three main categories: the economic census-related activities; the administrative data; and SBR capacity-building.

The following are some of the activities related to SBR during 2016:

> ‣ Checking the duplication of SBR data by SMA
> ‣ Updating the area master
> ‣ Updating the KBLI (Indonesian version of ISIC) master, from the years 2009–2015
> ‣ Producing sampling frames for businesses
> ‣ Preparing prefilled questionnaires for the 2016 economic census based on SBR data

In 2016, the administrative data sources were focused on the corporate sector, with the result that the study was limited to ministries and agencies that had administrative data sources of corporations. These include the Investment Coordinating Board (BKPM), the Social Security Administering Agency, Bank Indonesia, the Ministry of State-owned Enterprises, the Ministry of Trade, the Directorate General of Taxa-

tion, the State Electricity Company, YellowPages Indonesia, the General Directorate of General Law Administration in the Ministry of Law and Human Rights, the Ministry of Cooperatives and Micro Industry, small and medium-sized enterprises (SMEs,) the Indonesian Chamber of Commerce , and the Financial Services Authority. Prior to the visits to these institutions, profiling was first conducted to ensure that the SBR team understood the initial description of the institution of the administrative data source to be visited. Profiling was done through the collection of information from the official website of the ministries or institutions.

As a next step, the SBR team visited the ministries and institutions for validation and confirmation of the information that had been obtained from profiling. In addition to confirming information obtained through the Internet, visits were made to build good relationships with administrative data providers. During the visit, the SBR team presented SBR and administrative data studies to the ministries and institutions. Then the ministries and institutions presented the administrative data that they managed.

Based on the results of the study, three administrative data sources were chosen as the most promising primary administrative data sources: the Directorate General of Taxation, BPJS Employment and the State Electricity Company.

Given the continuing important roles played by the SBR, it was necessary to involve the BPS offices in the provinces in the SBR development process. For that reason, it was deemed necessary to introduce the SBR to the regional BPS offices by conducting an SBR capacity-building exercise. The materials given in this training event included: introduction of the SBR; SBR and BPS transformation; large business unit profiling; SBR system; SBR and the 2016 economic census; and the economic statistical unit model.

In 2016 the SBR team also successfully approached the Investment Coordinating Board for a memorandum of understanding, representing the first step in obtaining administrative data for future SBR updating purposes.

Activities in 2017

The activities undertaken in 2017 were part of the effort to update the SBR in two ways, as laid out in the Guidelines on Statistical Business Registers: survey and census feedback and profiling. The process of loading the SBR, followed by matching activity, was performed first. The next step was profiling to update the description of the structure of the State-owned enterprises and large and complex enterprise groups. Lastly, a field check was also carried out to confirm the results of the data integration.

Loading was done internally in the SBR secretariat by the Directorate of Census and Survey Methodology Development. The data obtained from the 2016 economic census listing were the data of the prefilled questionnaires, which had originally been prefilled from the SBR, and the non-prefilled questionnaires. Those data were then loaded into the SBR database system. The non-prefilled data were filtered in accordance with their formal status. For example, the household enterprises were included if they had legal status or separate financial reports.

The matching of SBR data was performed on the 2016 economic census listing results data. The SBR data that were matched consisted of business data on large and medium-sized enterprises from non-prefilled questionnaires. For matching work to be efficient, fewer numbers were matched with the bigger number, in this case, SBR data were matched with 2016 economic census listing data. For every business in the SBR, the SBR system searched for the most similar 25 businesses based on the results of the 2016 economic census listing. If any of the 25 similar business was the same business as the SBR business, then the two businesses were said to match and operators merged

the data from both sources. Otherwise, the SBR business was viewed as not matching any of the businesses of the 2016 economic census listing, with the effect that operators added the business to the SBR as new SBR data.

For profiling, preparatory activities were conducted and output had already been obtained in the form of a list of State-owned enterprises and enterprise or enterprise groups that had been profiled by staff members starting from the last week of November 2017. The list of State-owned enterprises was derived from the 2015 annual report on the acceleration of the strategic role of State-owned enterprises. While the list of enterprise groups was derived from the Globe Asia 2017 magazine, the list of the largest enterprise group in Indonesia in 2017 was based on revenue.

The first stage was to complete the main profiles of the largest enterprise groups and State-owned enterprises where this activity was entirely performed by internal staff members. The focused variables were the name of enterprise or enterprise group, equity, owner, and the area of business activity, direct subsidiary (if less than a 50 per cent share) in Indonesia, website and headquarters office. The internal staff members, who were divided into several groups, did this work simultaneously with the help of search engines to obtain the profiles of the businesses.

The second stage was to create a more detailed profile for each enterprise or enterprise group. This work was also done by dividing the internal staff members into several groups. Each group was responsible for multiple enterprises or enterprise groups.

The next activity was the business field check, conducted by SBR secretariat staff members together with SBR interns in several provincial BPS offices. This activity was carried out primarily to check the businesses that had not matched in the previous matching activities. In this way, it was hoped to ascertain the reason why some SBR businesses had net been included in the 2016 economic census listing. Some business cannot be found at the SBR address for various reasons, such as their having closed down or moved.

Current SBR system

Currently, the SBR system is characterized by five main features. The first relates to its data integration, which includes the batch uploading tool and matching tool. The batch uploading tool provides a facility for the users to upload a large volume of data in a predefined Microsoft Excel format so that they can be validated first in the SBR server before the matching process. The matching tool (see fig. C6.2) facilitates efforts by the user to match the incoming units that have been uploaded to the units that are already in the SBR database to avoid duplications. This tool is employed in particular when the incoming units have no SBR ID. For each incoming data item, the system searches for the top 25 similar units in the SBR database. If none of the 25 units matches the incoming data, then the incoming data can be added to the SBR database as a new unit. When, however, the incoming data item matches one of the top 25 similar units (usually with the unit in the first rank), the operator does not need to add it to the SBR database. The operator can still add or update the data in the SBR database with the information from the incoming data.

Figure C6.2
Matching tool in the Indonesian SBR system (with fictional data)

Similar companies from search results (category: G)	Matched: 15467 Not Matched: 1 Skipped: 13990 Remained: 1496

New company   File: Swasta - Rekso Group.xlsx

1. Company Name: PT Rekso Nasional Food McDonald's Indonesia Restoran Cabang Lampung Tanjung Karang Lampung
2. Commercial name:
3. Address: JIRA Kartini No 21 Lampung
4. Village:
5. District
6. Regency: 71 (BANDAR LAMPUNG)
7. Province: 18 (LAMPUNG
8. Country: 138 (INDONESIA)
9. Category: I (Penyediaan Akomodasi Dan Penyediaan Makan Minum)
10. KBLI: 56101 (Restoran)
11. Statistical unit: Establishment

Company 29458 from 15487 | Detail | Skip | Not Match |

SBR Companies (Search Results)   Same Name: 0

1. Company name: MDS Central Plaza Lampung
2. Commercial name:
3. Address: Jl Kartini No 21 Tanjung Karang Bandar Lampung Lampung
4. Village:
5. District
6. Regency: 71 (BANDAR LAMPUNG)
7. Province: 18 (LAMPUNG)
8. Country: 138 (INDONESIA)
9. Category: G (Perdagangan Besar Dan Eceran; Reparasi Dan Perawatan Mobil Dan Sepeda Motor)
10. KBLI: 47191 (Perdagangan Eceran Berbagai Macam Barang Yang Utamanya Bukan Makanan, Minuman Atau Tembakau Di Toserba (Department Store))
11. Statistical unit: Establishment

Id: 550650   Rank 1 of 25 | Detail | << | >> | Keep | Replace | Edit |

The operator can decide whether an incoming data item matches one of the top 25 similar units. If none of the top 25 similar units matches the incoming data, the operator can add the incoming business to the SBR database.

The second feature is the profiling feature. It provides facilities to manage the relationships between units in the SBR (see fig. C6.3) and manages the variables of the units. The profiler can also add a new unit in this feature in the profiling activity.

Figure C6.3
Screenshot showing the relation between units in the SBR system (with fictional data)
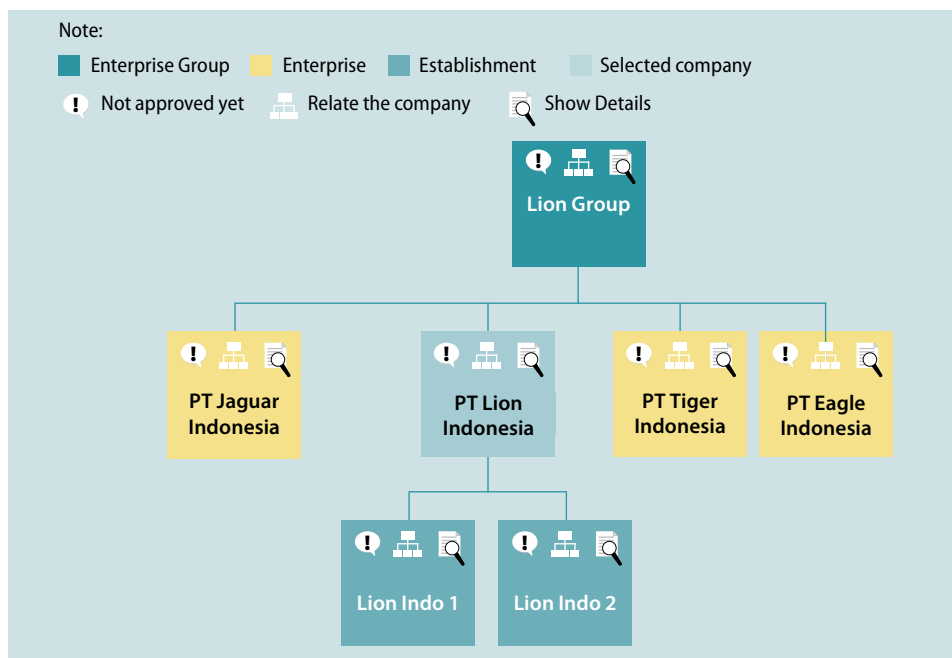
The third feature is the statistics feature. This provides a monitoring dashboard for each operator and also for the managers. The history of editing for each company can also be viewed here. With this feature, many activities of the SBR can be controlled.

The fourth feature is the browsing feature, by which people with access rights can see the profiles of a business or obtain a list of businesses based on some criterion or other, such industrial code, province, statistical unit types, and so forth.

The last feature for the current condition is the tutorial. Besides the training that users receive prior to starting their work, with this feature they can also learn about how to use the SBR system.

Annex D
Examples relating to statistical business register quality assurance

Annex D1: SBR quality indicators of Italy

D1.1: Introduction

The National Institute of Statistics of Italy (Istat) maintains an SBR, referred to as the Business Register of Active Enterprises (ASIA). Based on the principle of transparency, Istat produces a declaration of ASIA quality to accompany ASIA, viewed as a micro-data file. The declaration comprises a set of indicators intended to measure the various quality components. It is a synthesis of indicators, direct and indirect, that give a temporal reference to the data, sources and variables. It is easy to read and interpret in terms of users' needs. It covers several aspects:

‣ Use of metadata contained in the archive's database

‣ Compilation of indicators, presented as tables, graphs, histograms and in other forms

‣ Calculation of synthetic indicators

‣ Selection of priority indicators

The following paragraphs describe 50 quality indicators that can be calculated for each reference year, including the various breakdowns of these indicators by selected characteristics. They cover the basic data from the administrative files that provide the inputs to ASIA, as well as the statistical data on the output side, namely, the ASIA enterprises.

Comparisons over time (two or more consecutive years) and interpretation of increases and decreases in the indicators provide additional information on quality.

In the following paragraphs:

‣ t = reference year of ASIA data

‣ s = administrative source of input

‣ Chamber of Commerce (CCIAA)

‣ Fiscal register (MEF)

‣ Social Security (INPS)

‣ st = data from source s for reference period

D1.2: Quality of inputs

Quality criterion: timeliness

The indicators measure the time lag in the delivery of the administrative data to the SBR for each source. The time lag is the difference between the date on which the data are supplied and the reference period to which they refer.

1. Temporal lag, measured in months, between supply date (s_t) and reference year t, compiled for s_t = CCIAA.

2. Temporal lag, measured in months, between supply date (s_t = MEF) and reference year t, compiled for s_t = MEF.

Quality criterion: coverage

3. Number of records received from supply s.

Measurement of the completeness of enterprise births and deaths is useful in detecting undercoverage and overcoverage. It is better to use comparisons of data from a single source over time since simple counts of the numbers of births or deaths during a reference period do not of themselves provide much information.

The effect of update delay of the source is obtained by comparing the values of a characteristic – the number of cessations in year t (Ncess as in indicator 4) or the number of starts of activity in year t (Nstart as in indicator 5) – in the yearly supplies received at time t (s_t) and at time $t+1$ (s_{t+1}) – each case referred to the year t.

The two simple indicators to estimate the lag in the registration of the two dates in input source s = CCIAA:

4. Loss of information about cessations occurring in year t using the supply s_t:

    $1\text{-Ncess}[S_{t+1}(t)] / \text{Ncess}[S_t(t)]$

    where

    t = year of cessation

    $\text{Ncess}[S_t(t)]$ = number of cessations occurring in year t for which data was received in year t from source s.

    $\text{Ncess}[S_{t+1}(t)]$ = number of cessations occurring in year t for which data was received in year $t+1$ from source s.

5. Loss of information about starts of activity occurring in year t using supply s_t:

    $1\text{-Nstart}[S_{t+1}(t)] / \text{Nstart}[S_t(t)]$

    where

    (t)=year of starting activity

    $\text{Nstart}[S_{t+1}(t)]$ = number of starts occurring in year t for which data was received in year t from source s.

    $\text{Nstart}[S_t(t)]$ = number of starts occurring in year t for which data was received in year t from source s.

Quality criterion: completeness

The indicators measure the completeness of characteristics, including the numbers of missing values and/or number of missing values as a proportion of the total number of values, for selected characteristics.

6. Company name, s = CCIAA: number and per cent of records with missing company name

7. Legal status, s = CCIAA: number and per cent of records with missing legal status.

8. Address(es), s = CCIAA: number and per cent of records with missing address(es).

9. Principal economic activity code, s = CCIAA: number and per cent of records with missing principal economic activity code.

10. Territorial (municipality, province) code, s = CCIAA: number and per cent of records with wrong territory code.

11. Economic activity code, s = MEF: number and per cent of records with missing economic activity code.

12. Registered head office abroad, s = MEF: number and per cent of records with missing information about registered head office abroad.

13. Employees, s = INPS: number (per cent) of records with zero employees.

14. Economic activity code with classification not up to date, s = MEF: number and per cent of records with classification code according to obsolete classification.

D1.3: Quality of processes

Quality criterion: coverage

Information from the first macrophase (Integration of administrative sources and identification of units)

Records coming from different sources that pertain to the same legal unit (i.e. identified by a common taxation identification number) are integrated in order to build up a cluster of records for the same enterprise. The taxation register is, typically, the pivotal source, i.e. the base used to define the set of legal units and to integrate all the other sources.

15. No match: number and per cent of records in source s ($\neq$ MEF) not matched with s = MEF by tax code.

16. Undercoverage due to lag: number of clusters of administrative records indicating units not matched with MEF(t-1)* but then matched with MEF(t) and as per cent compared with number of *clusters* in MEF(t-1).

(*These are units that could have been linked and would have been included in the SBR in time (t-1) if the tax register had successfully updated such units in time.)

17. Structure of clusters by number of matching sources: Number and per cent of clusters consisting of 1, 2, … n sources.

Quality criterion: accuracy

Information from the first macrophase (Integration of administrative sources and identification of units)

18. Economic activity code (NACE): number of records for s = MEF with an old NACE classification, not coded using the new NACE codes as proportion of the total number of records with an old NACE classification.

Information from the second macrophase (Estimation of characters)

19. Activity status: number of units (and/or related employees) with activity status estimated using administrative sources, then modified by deterministic (automatic) rules.

Information from the third macrophase (Integration with statistical sources and quality control)

20. Number and per cent of units with erroneous NACE or activity status or number of employees.

21. Number of units (per cent of total errors) with erroneous NACE or activity status or number of employees then automatically corrected by standard deterministic rules.

22. Number of units (per cent of total errors) with erroneous NACE or activity status or number of employees then corrected manually by skilled BR staff (using clerical checks and online updating).

23. Number of units (per cent of total errors) with NACE or activity status or number of employees still to be ascertained and then accepted.

D1.4: Quality of outputs

The following indicators are calculated by size class in terms of number of employees in order to give different priorities to corrections.

Quality criterion: coverage

24. Number of enterprises active in *t* and as change from *t-1*.

25. Number of reactivations in *t* as a percentage of active enterprises.

26. Number of units (in terms of enterprises and employees) by source (survey, estimation, profiling) by characteristic (NACE, employees and status of activity).

Quality criterion: completeness

27. Incomplete NACE codes: Number of units with NACE 2-digit code; Number of units with NACE 3-digit code; Number of units with NACE 4-digit code).

28. Company name: number of units with company name missing.

29. Address: number of units with address missing.

30. Postal code: number of units with postal code missing.

31. Telephone: number of units with telephone number (fax, email) missing.

Quality criterion: timeliness

32. Temporal lag, measured in months, between the dissemination date of ASIA (i.e. date when ASIA data are made available to users) and the reference year to which they refer.

33. Latest information (with reference date later than *t* when disseminating data with reference year *t*): number of start dates later than *t*; number of cessation dates later than *t*; number of events (by type) having starting date later than *t*; number of units with employees updated at the time later than *t*.

Quality criterion: Accuracy

Direct indicators

Comparison between SBR data and SBS SME survey data concerning variables that are pre-printed in SBS questionnaires using SBR data:

34. Number and per cent of questionnaires rejected by type of error.

35. Number and per cent of units with wrong address (trend over the last x years).

36. Number and per cent of units with erroneous activity status (trend over the last x years).

37. Number and per cent of units with conflicting economic activity.

Comparison between SBR data and survey on the local units of big enterprises (concerning variables that are pre-printed in questionnaires for those units using SBR data):

38. Number and per cent of questionnaires rejected.

39. Number and per cent of units with modified NACE.

40. Number and per cent of units with modified and corrected NACE (by manual check).

41. Number and per cent of units with modified and erroneous NACE (by manual check).

42. Number and per cent of units with inconsistent employees.

43. Number and per cent of units with inconsistent and corrected employees (by manual check).

44. Number and per cent of units with inconsistent and erroneous employees (by manual check).

45. Number and per cent of units with modified activity status.

46. Number and per cent of units with modified and accepted employees (by manual check).

47. Number and per cent of units with modified and erroneous employees (by manual check).

Indirect indicators:

48. Status of activity: analyses of time series of active enterprises of ASIA (*t*, *t-1* and *t-2*) and measure of the reliability of selected subpopulations (defined by items like registrations, register deletions, reactivations, etc.). The indicator is also calculated by cell, like sector of economic activity and region.

Formula:

$$I_j = 100 - \left[\sum_k (x_{kj} \cdot \mathrm{val}_k) \, / \, \sum_k (x_{kj}) \cdot 100) \right]$$

where:

x = number of units related to each item (i.e. number of register creations).

k = selected subpopulation (item).

j = cell (class of employee, sector, etc.).

val = subjective measure assigned by statistician (BR staff) according to the importance of each item.

Mixed quality criterion: coverage, up-to-dateness, transparency, dissemination

50. Updates and changes referring to time *t* when disseminating *t+1* data: number (per cent) of false active units (non-active) in *t*; number (per cent) of units with different NACE in *t*; number of units with revised number of employees in *t* and related employees (per cent).

Annex D2: Experience of Colombia in implementation of quality processes

D2.1: Introduction

In Colombia, the process of updating the SBR, which is called the Statistical Directory, is supported by an information system called Directory Information System (SID).[77] The objective of the development of the SID is to automate processes and generate rules and controls enabling improvement in the quality of the information.

The SID is composed of eight modules, namely "supplier management", "information preparation", "information processing", "updating operations", "quality control", "data output", "management and configuration" "management and quality indicators". Although there are quality control processes throughout the modules, the article is focused on the information preparation and processing modules, because these are where rules are implemented to ensure that internal and external information is of the best possible quality and can be used for updating.

The main advances and results obtained by implementing a set of validation and standardization rules, and the process followed to combine administrative information with the SBR information base are described. Lastly, six indicators for measuring the quality of the SBR information are proposed.

D2.2: Description of the business register of Colombia

The SID aims to maintain an updated SBR by using information coming from administrative sources, updating operations (telephonic, field, and web) and statistical operations.

The process of updating the SBR covers three main components, namely, universe, coverage and traceability of statistical units. These components are described as follows.

- ➤ Universe: the SBR must contain all the enterprises (and corresponding legal and local units) conducting economic activities in the country.

- ➤ Coverage: the SBR must guarantee a national coverage of all the statistical units addressing economic activities in every sector, this latter determined by means of the Spanish version of ISIC, Rev.4, adapted for Colombia (CIIU Rev. 4 A.C.).[78]

- ➤ Statistical units: the conceptual data model refers to three basic units: the enterprise, the homogeneous production unit and the establishment. Nonetheless, the model is being adapted to a more general framework in which the legal units, enterprises, local units and enterprise groups are taken into account.

- ➤ Enterprise: economic entity or combination of the economic entities capable, by its own right, of possessing actives, incurring obligations and conducting economic and productive activities with other entities to achieve the objectives for which it was created.

- ➤ Homogenous production unit: characterized by a single activity: product inputs, production processes and homogeneous product outputs.

- ➤ Establishment: an enterprise or an enterprise part located in a topographically delimited place in which, or from which, economic activities

are conducted. Its implementation includes a set of variables classified in four categories: identification, location, stratification and management.

The traceability of the statistical units is made possible via the identification, location, stratification and management information, which is updated for each enterprise through the SID:

> ‣ Identification: the tax identification number (NIT),[79] social reason, acronym, commercial (trade) name, legal representative and juridical form are recorded, in addition to the identification keys for every statistical unit in the database.

> ‣ Location: department, municipality, address, web page, email, telephone.

> ‣ Stratification: CIIU code, busy staff, incomes from operational or sales.

> ‣ Management: constitution date, status, activity initiation and cessation dates.

D2.3: Quality improvement of the administrative information

The quality of the administrative information is measured in terms of coverage, reliability, coherence, opportunity, accessibility and traceability in the SBR frame. The updating and maintenance process is based on the following procedures:

> ‣ Processing and verification of administrative data.

> ‣ Feedback from economic surveys.

> ‣ Processing of economic unit data from censuses that are stored in the National Administrative Department of Statistics (which is a database).

> ‣ Own processes for verifying and validating the SID information.

The six major modules of the SID that guarantee the quality of the business register are: "information management", "information preparation", "information processing", "updating operations", "quality control" and "data output".

Information management

The information management module monitors information from the moment that it is requested from suppliers until it is received by the National Administrative Department of Statistics. The information received is checked to see whether it satisfies the minimum variable requirements to be used in the SID. The information is requested in a specific format. Definitions are provided.

A process includes a general review of every database provided, and the statistical record is generated containing a diagnosis of the information provided to support the preparation and loading processes.

Information preparation

The information preparation module consolidates the data provided by the suppliers at the level of the standard economic units (enterprise, establishment, and homogeneous production unit). A consolidated file with data from the different administrative sources is obtained. The variables in the supplier consolidated file are semi-automatically aligned with those of the SID.

[79] The tax identification number (NIT) constitutes the identification number of those enrolled in the Tax Unified Register (RUT), making it possible to individualize the contributors and users, for every effect on tax, custom and change matters, and especially for accomplishing such obligations.

Information processing

The information processing module automatically standardizes, codifies and applies rules of validation and consistency of the information consolidated from the suppliers to update the consolidated file. The following basic elements are distinguished in this processing:

- Validation rules incorporated to the SID.
- Unique identifier for each statistical unit.
- Address standardization for every statistical unit.
- Detection and elimination of duplicates.
- Creation of a catalogue to standardize words in names; for example, the word "Limitada" ("Limited") could be abbreviated as Lta, Ltda, Limit, etc., but the process standardizes it as "Ltda".

Some statistical units are sent to a base for revision when they do not satisfy the parameters. In addition, the information received from administrative sources is checked in this module.

The SID determines in automatic manner whether the statistical unit exists in the database and identifies the new units and the continuing units.

The reference date and origin of the value of every variable for every statistical unit is registered as the information is updated, so that every value is traceable.

Updating operations

Information updating operations are conducted by means of call centres, the Internet and electronic forms for economic units. The call centre technology platform has been upgraded to improve quality. The web form allows the enterprises to supply their information by directly electronically.

Quality control

There is a set of tools that enables analysis of any kind of information from administrative reports or contained in the SID.

Data output

A real-time consultation module is created that enables the generation of frames for the SID users.

D2.4: Results of the information preparation and processing processes

The information preparation module yields two main products: a consolidated base and a statistical report structured in four parts:

- Variable list with descriptions, provided by the supplier (dictionary).
- Attachment to the dictionary in which the variable classifications are presented.
- Basic diagnosis indicating duplicate, empty and inconsistent cells.
- Frequencies for every categorical variable, including department, municipality and legal organization, among others.

Codification rules

The database information received in accordance with the reference tables is codified in this process. Codification applies to *municipality, department, economic activity, type of document* and *legal organization.* For example, there are cases in which the information does not come codified but named with the department or municipality where the economic unit locates. The process transforms these data into the official codes according to the Colombian Political-Administrative Division (DIVIPOLA).

Normalization rules

The system allows normalization of economic unit address (location), name, telephone number, by applying defined and updated rules, according to the word normalization catalogues. Symbols like # or – in the addresses, are eliminated within the normalization process.

Validation rules

The completeness and consistency of information is verified as follows.

  ‣ The length of the telephonic numbers must be 7 or 10 digits, without the inclusion of the city indicative. The datum type must be numeric.

  ‣ The address must have more than 4 characters.

  ‣ The commercial (trade) name must have more than 4 characters.

  ‣ If the statistical unit is classified as active, and comes from a National Administrative Department of Statistics survey, it must have information about incomes and employed persons.

  ‣ Every statistical unit must have an identification unique number, which must be within a specific length range depending on the type of the associated document.

  ‣ The emails must have the symbol @ and belong to a valid domain.

D2.5: Proposed quality indicators

In this moment, Colombia is conducting the process of defining a set of indicators whose purpose is to evaluate the quality of SBR. The indicators proposed are as follows.

Indicator 1

Name: Updating level

Objective: To know the updating rate for every economic sector in the frame

Type of indicator: Process quality.

Variables used in the calculations are:

  ‣ Aj: Total updated records for sector *j*.

  ‣ Bj: Total records expected for updating in sector *j*.

The formula used for the calculation is:

$$I_{1j} = \frac{A_j}{B_j} * 100$$

Calculation frequency: Annual

Tolerance ranges:

> ▶   Critical ≤ 70
> ▶   70 < Fair ≤ 90
> ▶   Satisfactory > 90

Indicator 2

Name: Birth and death tracking

Objective: To evaluate the sector dynamics based on economic unit demographics

Type of indicator: process quality

Variables used in the calculations are:

> ▶ A_i: births in year i
> ▶ B_i: deaths in year i

The formulae used for the calculation are:

$$I_{2A}= \frac{A_i}{A_{i-1}}; I_{2B}=\frac{B_i}{B_{i-1}}$$

Calculation frequency: annual

Indicator 3

Name: Coverage.

Objective: to establish the SBR coverage compared to the information from the Tax and Customs National Directory (DIAN) database.

Type of indicator: Process quality

Variables used in the calculations are:

> ▶ t_i: Total records in the database in year i
> ▶ T_i: Total unique records identified in the national tax register database in year i

The formula used for the calculation is:

$$I_3= \frac{t_i}{T_i}* 100$$

Calculation frequency: Annual.

Tolerance ranges:

> ▶   Critical ≤ 70
> ▶   70 < Fair ≤ 90
> ▶   Satisfactory > 90

Indicator 4

Name: Employment precision

Objective: To determine whether the information by sector kept in the SBR is approximated to the official employment statistics generated by any National Statistics Institute

Type of indicator: Quality of the process

Variables used in the calculations are:

- ➤ e_i: total employees according to information in SBR in year i
- ➤ PEA_i: economically active population in the year i, according to official data
- ➤ L_i: lower limit established for estimating the unemployment by the national statistical office
- ➤ L_s: upper limit established for estimating the unemployment by the national statistical office

The formula used for the calculation is:

$$I_4 = \begin{cases} 1 & if\ L_i \leq 1\ \dfrac{e_i}{PEA_i} \leq L_s \\ 0 & otherwise \end{cases}$$

Calculation frequency: Annual

Tolerance ranges:

- ➤ 1 Satisfactory
- ➤ 0 critical

Indicator 5

Name: Income precision

Objective: To determine whether the information by sector kept in SBR is approximated to the official income statistics generated by any Institute of Statistics

Type of indicator: Quality of the process

Variables used in the calculations are:

- ➤ c_i: Total income according to information in the SBR in year *i*
- ➤ C_i: Total income according to official data of national accounts in year *i*

The formula used for the calculation is:

$$I_5 = \begin{cases} 1 & if\ 0.9_i \leq\ \dfrac{c_i}{C_i} \leq 1.1_s \\ 0 & otherwise \end{cases}$$

Calculation frequency: Annual

Tolerance ranges:

- ➤ 1 Satisfactory
- ➤ 0 critical

Indicator 6

Name: Opportunity

Objective: To establish the level in which registers are available for being used by SBR users

Type of indicator: Output quality

Variables used in the calculations are:

- ➤ e_i: Number of users qualifying the access as timely to SBR (schedule)

> ▸ *U*: Total number of interviewees

The formula used for the calculation is:

$$I_6 = \frac{u_j}{U} * 100$$

Calculation frequency: Annually

Tolerance ranges:

> ▸ Critical ≤ 70
>
> ▸ 70 < Fair ≤ 90
>
> ▸ Satisfactory > 90

Annex D3: Administrative data source evaluation checklist by CBS

Quality framework for administrative data sources

National statistical offices are increasingly making use of secondary data sources such as administrative registers for the production of statistics. However, the information in these sources is often collected and maintained by other organizations, usually for non-statistical purposes. Since the production of high-quality statistics depends on the quality of the input data, it is useful to evaluate the quality of secondary data sources, in a systematic, objective, and standardized way. For this purpose, CBS has developed a quality framework focusing on the quality of administrative and other secondary data sources.[80]

The framework distinguishes between the quality of the *data source*, the *metadata* and the *data*. For each of these the framework provides a number of dimensions and indicators that can be used to evaluate whether data from specific administrative or other secondary data are suitable to be used in the compilation of statistics. The details of the framework are provided in figures D3.1, D3.2 and D3.3.

The framework is applied by successively evaluating the quality indicators of the data source, metadata, and data. If a quality indicator of the data source reveals a problem, this should be addressed first, before starting to evaluate the indicators for the metadata. Likewise, any problems with the quality indicators of metadata should be addressed before evaluating the quality indicators of the data. If the evaluation of the quality indicators for all three areas is successful, the data source may be used for the production of statistics.

The framework includes detailed checklists for data source and metadata to be completed by the prospective internal user of an administrative or other secondary data

[80] Piet Daas, Saskia Ossen, Rachel Vis-Visschers and Judit Arends-Tóth, "Checklist for the quality evaluation of administrative data sources", Discussion Paper 09042 (CBS, 2009). Available at https://ec.europa.eu/eurostat/documents/64157/4374310/45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf.

source and/or an expert from the data source. For the data source checklist, it is advised to contact the official national statistical office contact person for the particular source (if there is one).

Figure D3.1
Quality aspects related to the data source

Source		
Dimension	**Quality indicators**	**Methods**
Supplier	1.1 Contact	Name of the data source. Data source contact information. National statistical office contact person.
	1.2 Purpose	Reason for use of the data source by the national statistical office.
Relevance	2.1 Usefulness	Importance of data source to the national statistical office.
	2.2 Envisaged use	Potential statistical use of data source.
	2.3 Information demand	Does the data source satisfy information demand?
	2.4 Response burden	Effect of data source use on response burden.
Privacy and security	3.1 Legal provision	Basis for existence of data source.
	3.2 Confidentiality	Does the Personal Data Protection Act apply? Has use of data source been reported by the national statistical office?
	3.3 Security	Manner in which the data source is sent to the national statistical office Are security measures (hard/software) required?
Delivery	4.1 Costs	Costs of using the data source.
	4.2 Arrangements	Are the terms of delivery documented? Frequency of delivery.
	4.3 Punctuality	How punctually can the data source be delivered? Rate at which exceptions are reported. Rate at which data is stored by data custodian.
	4.4 Format	Formats in which the data can be delivered.
	4.5 Selection	What data can be delivered? Does this comply with the requirements of the national statistical office?
Procedures	5.1 Data collection	Familiarity with the way the data are collected.
	5.2 Planned changes	Familiarity with planned changes to the data source. Ways changes are communicated to the national statistical office.
	5.3 Feedback	Can contact data custodian in case of trouble? In which cases and why?
	5.4 Fall-back scenario	Dependency risk incurred by the national statistical office. Emergency measures when data are not delivered according to arrangements made.

Figure D3.2
Quality aspects related to metadata related aspects of the data source

Metadata Dimension	Quality indicators	Methods
Clarity	1.1 Population unit definition	Clarity score of the definition.
	1.2 Classification variable	Clarity score of the definition.
	1.3 Count variable	Clarity score of the definition.
	1.4 Time dimensions	Clarity score of the definition.
	1.5 Definition changes	Familiarity with changes that occur.
Comparability	2.1 Population unit definition comparison	Comparability with the national statistical office definition.
	2.2 Classification variable definition comparison	Comparability with the national statistical office definition.
	2.3 Count variable definition comparison	Comparability with the national statistical office definition.
	2.4 Time differences	Comparability with the national statistical office reporting periods.
Unique keys	3.1 Identification keys	Presence of unique keys. Comparability with unique keys used by the national statistical office.
	3.2 Unique combinations of variables	Presence of useful combinations of variables.
Data treatment (by data custodian)	4.1 Checks	Population unit checks performed. Variable checks performed. Combinations of variables checked. Extreme value checks.
	4.2 Modifications	Familiarity with data modifications. Are modified values marked and how? Familiarity with default values used.
	4.3 Punctuality	How punctually can the data be delivered? Rate at which exceptions are reported. Rate at which data are stored by data custodian.
	4.4 Format	Formats in which the data can be delivered.
	4.5 Selection	What data can be delivered? Does this comply with the requirements of the national statistical office?

Figure D3.3
Quality aspects related to the accuracy of the data

Data Dimension	Quality indicators	Methods
Technical checks	1.1 Readability	Can all the data in the source be accessed?
	1.2 Metadata compliance	Does the data comply with the metadata definition? If not, report the anomalies.
Over coverage	2.1 Non-population units	Percentage of units not belonging to population.
Under coverage	3.1 Missing units	Percentage of units missing from the target population.
	3.2 Selectivity	R-index1) for unit composition.
	3.3 Effect on average	Maximum bias of average for core variable. Maximum RMSE2) of average for core variable.
Linkability	4.1 Linkable units	Percentage of units linked unambiguously.
	4.2 Mismatches	Percentage of units incorrectly linked.
	4.3 Selectivity	R-index for composition of units linked.
	4.4 Effect on average	Maximum bias of average for core variable. Maximum RMSE2) of average for core variable.
Unit non response	5.1 Units without data	Percentage of units with all data missing.
	5.2 Selectivity	R-index for unit composition.
	5.3 Effect on average	Maximum bias of average for core variable. Maximum RMSE of average for core variable.
Item non response	6.1 Missing values	Percentage of cells with missing values.
	6.2 Selectivity	R-index for variable composition.
	6.3 Effect on average	Maximum bias of average for variable. Maximum RMSE of average for variable.
Measurement	7.1 External check	Has an audit or parallel test been performed? Has the input procedure been tested?
	7.2 Incompatible records	Fraction of fields with violated edit rules.
	7.3 Measurement error	Size of the bias (relative measurement error).
Processing	8.1 Adjustments	Fraction of fields adjusted (edited).
	8.2 Imputation	Fraction of fields imputed.
	8.3 Outliers	Fraction of fields corrected for outliers.
Precision	9.1 Standard error	Mean square error for core variable.
Sensitivity	10.1 Missing values	Total percentage of empty cells.
	10.2 Selectivity	R-index for composition of totals.
	10.3 Effect on totals	Maximum bias of totals. Maximum RMSE of totals.

1) R-index: Representative Index, an indicator that estimates the selectivity of the data missing by using information available in other sources (Schouten and Cobben 2007, Cobben and Schouten 2008).
2) RMSE: root-mean-square error; statistical measure for the quality of an estimator. The RMSE is equal to the square root of the sum of the bias and variance of the estimator.

Annex E
Additional concepts and procedures

Annex E1: Register-based census

E1.1: General description

An economic census provides information about the structure and function of a production system from the national (macro area) to the local (micro area) level. In general, it guarantees periodic and comprehensive statistics about businesses, establishments, activities carried out and employment every 5 or 10 years. Economic census statistics are essential for economic policy and business planning. They provide an essential framework for all economic indicators (production indexes, input/output measures, labour, etc.) and they are fundamental inputs for benchmarking GDP estimates.

Basic statistical data on businesses are important and necessary for the public and private sector in policy formulation and development planning of the economy and industry at both national and local level. National and local governments use these data to monitor economic activity and the changes in national and regional economies. Census data are very useful for the individual businesses in calculating market share, locating business markets, identifying business site locations, and evaluating new business opportunities.

In many countries the censuses are regulated by law and provide for mandatory responses. In recent decades, the development, in a wide range of countries, of the national SBR guarantees a significant improvement in methods and tools for census data collection.

At one end of the spectrum is the traditional census, in which data are collected by use of enumerators and questionnaires, using no register information at all. At the other end is the totally register-based census. Some countries use mixed mode data collection with a combination of data from registers and questionnaires (either total enumeration or a sample survey). Even countries conducting mainly traditional censuses tend to use register information to some extent, for instance as an address list.

The interaction between the SBR and the census can be defined as "register-assisted census data collection" that combines some elements of a direct door-to-door survey and some of a classic survey by list. This technique of data collection is characterized by the following three elements:

> ‣ Enumerators are supplied with lists of the enumeration units located in their census districts, drawn from the SBR. Their task consists in verifying the actual status of the listed units, deleting the records of the duplicates and of those that ceased to exist, and adding new records for the possible non-listed units (for example those born in the lag between reference period of the list and the date of the survey, or unregistered for any other reason).

> ➤ Some days before the survey, all the listed units receive by mail a personalized questionnaire partly completed with information drawn from the SBR. In this way, the respondents just have to verify the correctness of the pre-printed fields (rectifying them, if necessary) and complete the questionnaire with the missing information.

> ➤ The questionnaires are collected by enumerators. The enumerators are provided with blank (non-personalized) questionnaires to be used only for non-listed units or in substitution of personalized questionnaires that are lost or damaged.

In some countries, where the mail service is not so efficient, the enumerator applies a prefilled questionnaire, using a computing mobile device or a printed questionnaire, when updating the list of units.

Using this method, based on the synergy between the census and the SBR, the main benefits are:

> ➤ *Reduction in burden for respondents*, since some fields of the questionnaire are prefilled, and just need to be updated by the respondent; this simplifies the procedure and reduces the time required for data collection and processing, which in turn has a positive impact on the quality of the data .

> ➤ *A new approach to quality and coverage control*, since it is possible to carry out a micro-level coverage analysis, by comparing the raw data collection file with an image of the SBR as of the same date as the census. This micro approach, instead of the classical macro approach based on a post-enumeration sample survey, makes possible the precise identification of every single unit undercovered or overcovered in both data sources. The results are that the theoretical undercoverage – the main problem of a direct door-to-door survey – and overcoverage – the main problem of an SBR – are significantly reduced.

> ➤ *Creation of a basis for improvement of the SBR* in terms of new characteristics obtained, for example, secondary activities.

> ➤ *From the results of the economic census*, the SBR is fully updated; accordingly, the economic census is also an updating source of the SBR.

An evolution of the previous register-assisted approach is the development of a register-based economic census. With this approach the census data are reproduced entirely by integrating data from the SBR and administrative sources, without any direct data collection from businesses. The potential for turning a traditional business census to a register-based one depends mainly upon the degree of enhancement achieved in the statistical use of administrative sources and on quality improvements by each administrative body in the business area.

The advantages of the register-based over a register-assisted approach are:

> ➤ A significant decrease in costs for the statistical authorities.

> ➤ The non-existence of respondents means the absence of respondent burden.

> ➤ Statistics can be made available every year, providing an opportunity to detect shorter-term changes in the economic structures of the national and local economies.

On the other hand, besides the classical problems that arise in the use of administrative sources (the need to deal with administrative definitions and their operational

rules, the timeliness in the production of data, the treatment and the exploitation of new administrative data, the enlargement of the data set of information, the huge amounts of data linked), the key disadvantage of this approach is the absence of direct statistical information to improve SBR quality, especially in terms of undercoverage and overcoverage.

There are some significant criteria for implementation of a register-based census:

> ▸ *High-quality SBR in terms of coverage.* The population, in terms of enterprises and local units, is decided and cannot be changed or corrected.

> ▸ *Existence of a unified identification coding system across different administrative sources.* In the absence of such unified systems it is extremely difficult, if not impossible, to link different registers. Integration is of key importance to a register-based census.

> ▸ *Methodology for translating administrative characteristics into statistical characteristics.* There is need to develop appropriate statistical methodologies (probabilistic and/or deterministic) to translate or to estimate statistical characteristics starting from one (or more, linked) administrative data sources.

The realization of a register-based census depends on having the appropriate national statistical objectives and sufficient national administrative sources to support the objectives. There is no single recipe. The actual approach depends upon the country situation, including the availability and completeness of administrative sources.

E1.2: Italian virtual economic census

For the first time the decennial business census, referred to as "CIS 2011"*,* aiming at the enumeration of businesses, related persons employed and other types of employment, classified by activity code, size, juridical status and other structural information about enterprises, was carried out without any direct data collection from business. Instead, it was based exclusively on statistical data obtained from the integration of administrative sources.

The virtual economic census system is built around a set of basic registers containing comprehensive data on business units and individuals. The core of this system is the national SBR (ASIA BR), which is produced yearly through the integration of administrative and statistical sources. ASIA BR is considered the reference population and the official source of Italian statistical information on the structure and the demography of the business population.

In order to produce census data for enterprises that are active at the end of December 2011 (*t*), the ASIA BR uses information available in the period from *t*+6 months to *t*+12 *months* from a set of administrative sources. On this basis the census data can be disseminated at *t+18* months, which represents a reduction of the time lag compared to the dissemination of previous traditional censuses results.

The identification and acquisition of a new set of administrative sources changed the way in which the process is carried out, and added new content to the database of business units and individuals, in which information is available not only at unit (enterprise, local unit) level but also at an individual level. Each person involved in the business production process is identified. In fact, each unit inside an administrative source is linked to the ASIA BR statistical unit by means of identification codes. Persons can be linked to the business in which they assume any ownership share, to the employer for whom they are working, and so on.

The virtual economic census makes use of many administrative sources. They may be grouped into different types:

- ▸ Fiscal data: VAT, income, participation in partnerships, remuneration taken from the 770-Form tax register, or the statistics-based tax assessment fiscal survey – the so-called "Studi di settore".

- ▸ Social security data: monthly employer declarations on each employee, on outworkers – namely, persons that are linked to a production unit and that are not employer or own account workers or employees – and on self-employed workers and family workers in agriculture, handicraft and trade.

- ▸ Chamber of Commerce data: list of partners or shareholders of each legal unit.

- ▸ Employment insurance data: from the government agency responsible for insurance against work-related injuries (workplace, insurance payments).

The integration of these sources has been made possible by the existence of a tax identification code that is available – with a very high level of coverage – both for legal units and individuals, in all sources and in ASIA BR. Thus, in Italy, the tax code represents an indirect and unified identification code system.

Using information from these sources, a new employment data structure for the SBR was set up – the linked employer-employment database (LEED) structure. While in the old SBR the number of persons employed represented one of the characteristics associated with a statistical unit (attribute), in the new data structure each single person is linked to each statistical unit for which he or she has any form of employment (according to the conceptual framework of the employment classification). In this new system, businesses and workers employed are identified by links derived by the integration and matching processes for the various administrative sources. The data structure is based on links, such as direct relationships between employment identification codes and enterprises, together with the basic enterprise attributes, employment composition at enterprise level (e.g. gender and age composition and workplace) and employment attributes that differ according to the type of employment.

This new integrated approach is able to provide an in-depth analysis of the employment of the enterprises and of local units. This analysis was, and still is, the main objective of the Italian virtual economic census. The main outputs of this census are the following:

- ▸ Data on employment, with a particular focus on the demographic characteristics of the workers, such as gender, age, birth location (country).

- ▸ More detailed information and job characteristics (professional status, type of contract, whether full or part time, and so forth).

- ▸ Identification of the different types of workers used in a business (employees, self-employment, family worker, outworkers, temporary workers) to provide a global picture of the labour input, both for each enterprise, and at sectoral and territorial levels.

- ▸ Identification of and solutions to measurement issues (hours actually worked, number of jobs, FTE jobs, number of persons employed).

From conceptual point of view the change in moving to the virtual economic census system may be summarized in the two schemas in figures E1.1 and E1.2.

Furthermore, the virtual economic census provided an opportunity to face and solve problems with employment definitions, in terms of improving coherence with the international standards and translating such definitions into operational rules. The

global revision of the employment classification system and of employment measurement methods has been a fundamental outcome of the virtual economic census, providing improved coherence for the whole national statistical system.

Figure E1.1
Old conceptual approach: SBR – Italy (ASIA)

| SBR | ←→ | SBR unit characteristics |

- Legal unit
- Enterprise
- Local unit
- Enterprise group

- Identification characteristics (ID, name, address, legal status)
- Demographic characteristics
- Economic stratification characteristics (activity code, number of employees, number self-employed, turnover)
- Control/ownership relations
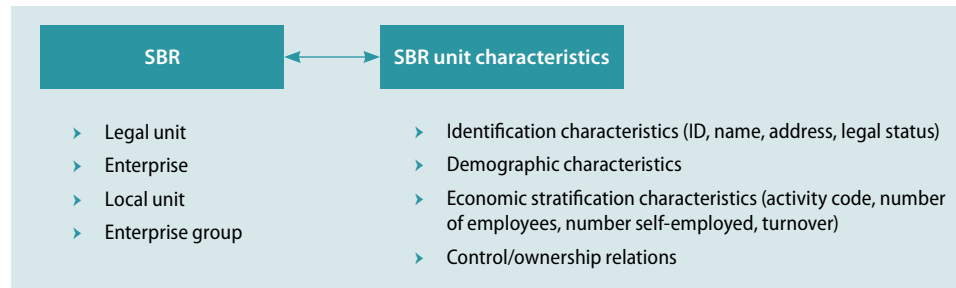
Figure E1.2
New conceptual approach: Integrated system for Virtual Census

SBR units	←→	SBR unit characteristics
Work relationship	←→	Work-related characteristics
Worker	←→	Individual characteristics
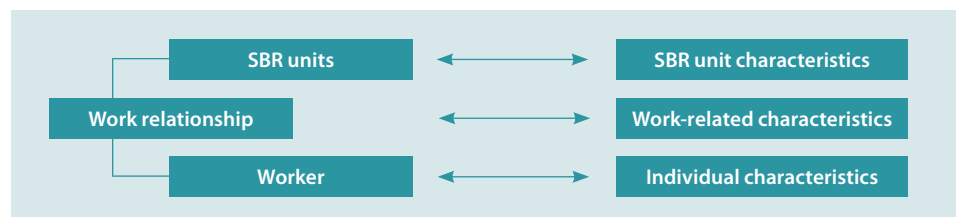
Annex E2: Calculating a check digit for an identification number

Introduction

A "check digit" is a decimal (or alphanumeric) digit added to an identification number. It is a form of redundancy check that assists in detection of errors in identification numbers (e.g. bank account numbers) that have been manually entered. It is analogous to a binary parity bit used to check for errors in computer-generated data. It is computed by an algorithm from the other digits (or letters) in the identification number.[81]

The algorithms used to generate check digits are designed to detect typical human transcription errors. In order of complexity, these include:

- Single-digit errors, such as 1 becoming 2
- Transposition errors, such as 12 becoming 21
- Twin errors, such as 11 becoming 22
- Jump-transposition errors, such as 132 becoming 231
- Jump-twin errors, such as 131 becoming 232
- Phonetic errors, such as 60 becoming 16

In choosing an algorithm, the high probability of detecting errors is traded off against implementation difficulty. Simple check digit algorithms are easily understood and

[81] Extracted from Wikipedia: http://en.wikipedia.org/wiki/Check_digit.

implemented by humans but do not detect as many errors as complex ones, which, however, require sophisticated programs to implement.

Use of check digits in the Swiss Business Register

In the Swiss Business Register, an algorithm based on Modulo 11[82] is used to calculate the check digits for all identification numbers. The characteristics of the algorithm are different from those based on other modules, such as the more common Modulo 10.

Local unit identification number (BURNR)

BURNR in the Swiss Business Register is an 8 digit number. The last digit is the check digit number.

> ➤   Example of BURNR: 62088168 – the final 8 is the check digit.

The procedure for calculating the check digit is as follows:

Each of the first 7 digits of the identification number is individually multiplied by a multiplier selected from a fixed sequence (5, 4, 3, 2, 7, 6, 5) according to the position of the digit within the number.

> ➤   The digit in position 1 is multiplied by 5
> ➤   The digit in position 2 is multiplied by 4
> ➤   The digit in position 3 is multiplied by 3
> ➤   The digit in position 4 is multiplied by 2
> ➤   The digit in position 5 is multiplied by 7
> ➤   The digit in position 6 is multiplied by 6
> ➤   The digit in position 1 is multiplied by 5

All resulting products are added and the result is divided by 11

> ➤   If the residual (Modulo 11) is zero the check digit is zero
> ➤   If the residual is 1, the original 7 digit number is not used as an identification number – it is considered invalid
> ➤   Otherwise the check digit is the number obtained by subtracting the residual from 11

Examples of calculation of check digits

Calculation of BURNR

BURNR	62088168
Effective digits	6208816
Sequence of multipliers	5432765
Products	(6*5=30) + (2*4=8) + (3*0=0) + (8*2=16) + (8*7=56) + (1*6=6) + (6*5=30) = 146
Result	146 /11 = 13 residual 3
Check digit	11 - 3 = 8

Enterprise identification number

A similar procedure is used for the enterprise identification number, which has nine digits including the check digit. The sequence of multipliers (5, 4, 3, 2, 7, 6, 5, 4). The first seven numbers are the same as for the BURNR.

Enterprise identification number	109322551
Effective digits	10932255
Sequence of multipliers	54327654
Products	(1*5=5) + (0*4=0) + (9*3=27) + (3*2=6) + (2*7=14) + (2*6=12) + (5*5=25) + (5*4=20) = 109
Result	109 /11 = 9 residual 10
Check digit	11 - 10 = 1

Unique enterprise identification number (UID)[83]

Exactly the same procedure is used for the unique enterprise identification number.

Unique enterprise identification number	CHE109322551
Effective digits	10932255
Sequence of multipliers	54327654
Products	(1*5=5) + (0*4=0) + (9*3=27) + (3*2=6) + (2*7=14) + (2*6=12) + (5*5=25) + (5*4=20) = 109
Result	109 / 11 = 9 residual 10
Check digit	11 - 10 = 1

[83] UID Register: www.uid.admin. ch/Search.aspx?lang=en.

Annex F
Methodology and country practice of profiling

Annex F1: European profiling

Introduction

European profiling is a collaborative method involving all the countries that host enterprises on their territory that are part of the same multinational enterprise group (MNE). The national statistical offices that take part in European profiling define and agree on the legal, financial and economic structure of groups and delineate their main economic activities irrespective of geographical borders. This common understanding of the groups' structure and their activities can strengthen data collection and data compilation and ensure more vertical and horizontal consistency in the European Union.

European profiling is based on a top-down approach. This means that profilers start their work from the group as a whole. From there they derive the constituent enterprises based on economic relevance, autonomy and potential to collect data.

There are a number of important differences between national and European profiling, in particular in terms of data sources. With national profiling, the national truncated part of an enterprise group is the starting point, as each country is able to observe only what is located in its national territory; with European profiling, the starting point is the analysis of the group irrespective of its geographical borders.

Using the top-down approach, European profiling aims to delineate enterprises within the enterprise groups. To delineate the enterprise, the starting point is the main economic activities of the group, the so-called "operational segment(s)", irrespective of any territorial border. The geographical dimension is added later in the process to clearly identify the national parts of each operational segment. This top-down approach and the resulting outcome can be a starting point for delineating the statistical unit enterprise, which is always carried out in accordance with Council Regulation (EEC) No. 696/93 and following the Operational Rules described in Chapter 7 – Statistical units of the European Business Statistics Manual.[84]

Different situations may arise in practice, which require adaptations to the outcome in order to fulfil the requirements of statisticians involved in data collection and data compilation at the national level.

European profiling implements the following units and terminology:

  ›   Global enterprise groups

An enterprise group can consist of several subgroups, which may be located in several countries. In European profiling terminology, the whole enterprise group is called a global enterprise group (GEG) to signify that it is considered an overarching unit due to its global nature.

[84] Eurostat, *European Business Statistics Manual*, 2021 edition. https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-21-001.

GEGs provide the overarching structure to which administrative legal units are linked.

> ► Global enterprise

According to the updated European profiling methodology, the global enterprise (GEN) is not to be considered a statistical unit in the European Statistical System. From an economic point of view, it is an autonomous unit that fulfils the concept of the enterprise irrespective of national boundaries but is not used in official statistics.

In practice, global enterprises often correspond to "operational segments" of an MNE, according to the International Financial Reporting Standards (IFRS) on the compilation of financial reports. In fact, operational segments can be used as the starting point for defining global enterprises.

The profiled global enterprise may differ from the IFRS operational segments used to manage the group for several reasons. Indeed, not all MNEs adopt IFRS or define their operational segment in terms that are relevant for statistical data collection. The operational segments are sometimes defined by geographical areas and therefore do not reflect the global enterprise concept and its activity, and cannot be used as such. In other cases, the operational segments may have no economic relevance for structural business statistics, no autonomy, or they are not suitable for statistical purposes. In all such cases, further analysis is necessary, and the collaboration of the groups is essential. Once identified, the global enterprise is a useful technical unit for supporting the delineation of the national enterprises that can be linked to it.

> ► Temporary enterprise

The updated European profiling methodology has introduced the concept of the temporary enterprise. This is the national part of a global enterprise proposed by the country of global decision centre to the respective country of residence of the enterprise. The latter is the only responsible for the final decision to validate it and recognize it as a statistical unit enterprise. The temporary enterprise is not a statistical unit in the European Statistical System. It is only a technical unit that can be automatically generated in the European profiling process and thus potentially used at national level as the starting point for delineating the statistical unit enterprise.

> ► Legal unit

According to the Regulation EU No. 696/93, legal units include: legal persons whose existence is recognized by law independently of the individuals or institutions which may own them or are members of them, and natural persons who are engaged in an economic activity in their own right. (EU No. 696/93, section II, A).

> ► National enterprise

The temporary enterprise might or might not correspond to the national enterprises as defined by Council Regulation (EEC) No. 696/93 and the Operational Rules for statistical units. The difference can consist of different economic activities and different sizes. When the temporary enterprise does not correspond to an enterprise, the national statistical institute involved in delineating the enterprise can modify its characteristics and size. The updated European profiling methodology considers different situations possible: the national enterprise can have a different economic activity with respect to the proposed temporary enterprise, at national level there may be several enterprises instead of one single temporary enterprise linked to a GEN or at national level there may be several temporary enterprises that need to be combined together to form a national enterprise. In all such situations it is the responsibility of the national statistical institutes involved in delineating the enterprise to make the necessary changes and assess the statistical unit enterprise that is suitable for data collection and for compiling data in a way at national level.

Profiling methods

There are different methods for profiling an enterprise group. These depend mainly on the size and complexity (number of legal units, their activities, and geographical locations) of the group, its economic relevance and the possibility to establish direct contact with group representatives.

As recommended by Eurostat on European business profiling,[85] profilers analyse all available information on a group and delineate enterprise(s) within this group on an individual basis. This method is costly in terms of time and resources and requires specific skills from the profilers, therefore, because of resource constraints, it tends to be applied to the largest, most relevant groups.

Intensive manual profiling takes place where there is a comprehensive dialogue and a meeting between the profiling team and the representatives of the multinational group.

Light manual profiling takes place where there is an analysis of the group without contact with its representatives, or contacts only by mails and conference calls with its representatives on very general aspects regarding global enterprise versus national enterprise (GEN/ENT) delineation.

Since a significant majority of groups in Europe are small and medium sized, profiling these groups is recommended to take place via an alternative method, the so-called automatic profiling. It refers to the automation process used to delineate small and medium-sized enterprises.

Annex F2: Profiling in Canada

In the Canadian Statistical Business Register (CSBR), the structure of a complex business encompasses (and ties together) two aspects: the legal structure and the operating structure, both of which are maintained and created by SBR enterprise profilers, using a variety of information sources and methods.

For the legal structure, all legal units are identified through tax data. The legal units in a common enterprise group are then linked to one another in a hierarchical structure representing majority ownership relationships.

For the operating structure, a profiler must identify the distinctly located physical places where the business engages labour and capital in an economic production and collect or derive the necessary register attributes for each of these "operating entities", including the industry code, number of employees and, when applicable, the operating revenue or turnover. Also critical, profilers must assess the manner in which the business accounts for the operating entities' activities, so that the CSBR system can classify it according to whether it is a statistical "establishment" or a statistical "location" or both. The distinction drives how Statistics Canada will target the collection of data from or about the business' activities. In essence, an establishment represents one or multiple statistical locations for which operating costs and operating revenue can be obtained.

The operating structure is then derived such that each location is linked to a "parent" establishment. Each establishment in turn is linked to the statistical company that controls it, thereby integrating the legal and operating structure of the business into a complete "complex business structure".

A feature of this approach is that profilers can verify the revenue and employment data they derive for the operating entities against the tax-data collected equivalent for the parent legal units. This is an important "coherency check" that helps bring consistency to the CSBR and the broader statistical programmes that it serves.

[85] *European business profiling: Recommendations manual,* 2020 edition. https://ec.europa.eu/eurostat/documents/3859598/10479728/KS-GQ-20-002-EN-N.pdf/e13f0907-5e5a-7521-604a-287004d07043.

For the System of National Accounts programme in Canada, it is important that the CSBR delineates business sectors and employment at the subprovincial level to calculate the GDP and employment across Canada. The CSBR must therefore delineate below establishment level to the location level when applicable.

The CSBR delineates each enterprise group or enterprise structure into four statistical units which correspond to the United Nations Guidelines on the Statistical Business Register as shown in figure F2.1.

Figure F2.1
Correspondence of statistical units in the CSBR to the United Nations

CSBR	United Nations
Enterprise	Enterprise
Company	
Establishment	Establishment Kind-of-activity unit
Location	Local Unit

The statistical units defined in the CSBR are described in figure F2.2.

Figure F2.2
Statistical units in the CSBR

Enterprise	‣ The organizational unit of a business that directs and controls the allocation of resources relating to its domestic operations, and from which international transactions, an international investment position and a consolidated financial position for the unit can be derived. ‣ It corresponds to the institutional unit as defined for the System of National Accounts.
Company	‣ The lowest level organizational unit for which income and expenditure accounts and balance sheets are maintained, from which operating profit and the rate of return on capital can be derived. ‣ Unit is part of a consolidation.
Establishment	‣ A unit for which the accounting data required to measure production are available (principal inputs, revenues, salaries and wages). ‣ The most homogeneous unit of production for which the business maintains accounting records. ‣ It is possible to assemble all the data elements required to compile gross value of production.
Location	‣ A production unit at a single geographical location at which or from which economic activity is conducted and for which, at a minimum, employment data are available. ‣ Either revenue or costs are available, but not both.

In the CSBR each statistical unit is assigned a statistical flag (represented by the blue boxes to the left in the following diagram) and a responsibility centre type as described in figure F2.3.

Figure F2.3
Responsibility centres
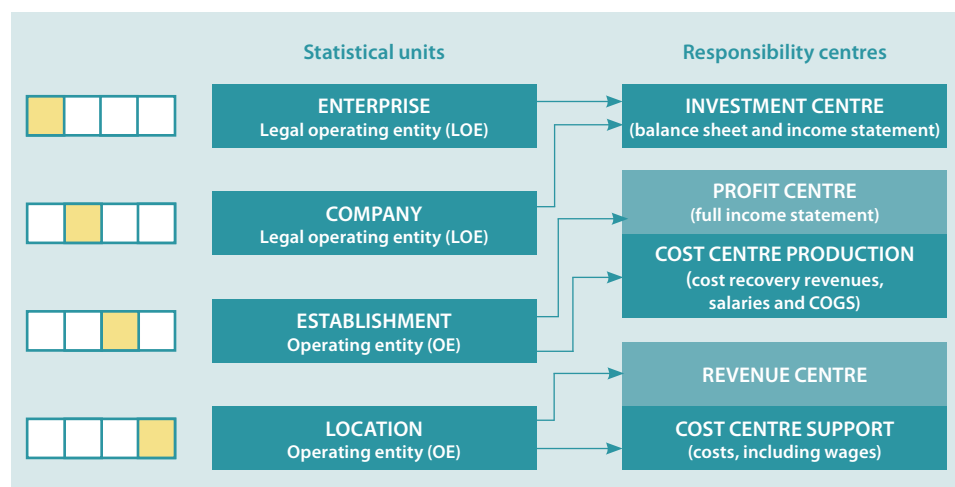
The responsibility centre type serves two purposes:

(1)   It uses terminology more relatable to business respondents during the profiling activity. They can better understand what it is that Statistics Canada is looking for while trying to represent the economic activity, the function and the reporting capability of the unit within the enterprise group or enterprise structure.

(2)   It clarifies the reporting capability of a unit to all economic programmes at Statistics Canada - most pertinent when looking at all establishments and locations as they can each be represented by two varying responsibility centres.

Note: This level of granularity is quite complex to maintain and is resource intensive. It should not be undertaken if not specifically required by the statistical programme of a country.

Prioritization of work

Considering limited resources to maintain the CSBR on an annual basis, a methodology has been developed to prioritize the large, high impact enterprises and enterprise group structures for profiling.

This methodology assesses the importance of each individual establishment based on:

‣   Assigned industry sector

‣   Geographical location

‣   Revenue, sales and turnover

‣   Employment numbers

The establishments identified as being of the highest importance are flagged as critical units on the CSBR. The enterprises that own and manage these establishments are prioritized for profiling activities on an annual basis.

Research tools

The profiling activity requires a significant amount of research to ensure the accuracy of the legal and operating ownership structure of an enterprise group and each enterprise within the group. For each establishment or location statistical unit in the structure, a profiler must assign the industrial activity, the reporting capability, the full civic address and the applicable size measures.

The following is a list of external information used throughout the profiling activity:

▸ Federal or Provincial Gazettes to determine reasons for legal status changes, industrial activity and legal address.

▸ Official company website to validate industrial activity, production locations, head office locations and company structure. Company websites can include financial statements, investor information, annual reports or budgets (if the company is non-profit).

▸ The System for Electronic Document Analysis and Retrieval (SEDAR) provides disclosure documents, financial statements, annual reports and company profiles for Canadian public companies.

▸ The Electronic Data Gathering, Analysis and Retrieval system (EDGAR) or other foreign governmental sources that shows if the enterprise has local units, subsidiaries, associates or parents in another country, search government sources if applicable for more information.

▸ General Internet search for recent news on the enterprise. Search for news on operating closures or births, legal bankruptcy filings, legal acquisitions, parent or subsidiary changes, etc.

The following is a list of internal information used throughout the profiling activity:

▸ Administrative data received from the Canada Revenue Agency through the business registration process for various tax programmes such as corporate, value added and payroll taxes.

▸ Business registration information providing the legal information for each registered company including industrial activity descriptions.

▸ Tax schedules providing detailed breakdowns of reported incomes and expenditures.

▸ Employment numbers for each registered business at the legal level.

▸ Business status information as reported to the Agency.

▸ Survey data from all programmes.

▸ Profilers have access to all monthly, quarterly and annual survey data collected for frame and coherence assessment purposes.

▸ Integrated collection systems allow regional office interviewers to provide feedback to profilers on changes to a business.

▸ Economic programme analysis and research information can be sent to the profilers.

▸ Historical information contained in the CSBR.

▸ Updates applied on the CSBR are recorded in a logging system which provides a history of the changes made and the update source.

▸ Journal reports entered into the CSBR provide background information on the updates processed or justification as to why updates were not processed.

Profiling phases and steps

Profile scope determination phase

Profilers have two options when determining what type of profile they must conduct.

Enterprise Profile – one legal unit – validate and update the direct parent and operating structure.

Enterprise Group Profile – group of related companies – validate and update the full legal relationship structure and the operating structure for each legal entity in the group.

Check the last time the enterprise group was profiled. If the last profile date exceeds three years, a full enterprise group profile is recommended.

Review and analyse all requests for updates sent through the CSBR system. Do they signal that changes are required to the entire enterprise group or to specific enterprises only? How many enterprises within the structure are affected?

Review the coherency of the structure. The CSBR has an integrated tool that can be used to quickly identify incoherencies in financial variables, employment numbers and vintage dates by comparing the operating entity data to the administrative data reported at the legal level.

Conduct an initial Internet search to identify other significant changes.

Using all or parts of the data collected in the above steps, determine the appropriate scope of the profile.

Data review and analytics phase

The amount of work involved in this stage will be determined by how current the CSBR data is, the size of the legal and operating structure, and availability of information on the structure being profiled. This phase can be time-consuming and resource-intensive. It requires access to all the research tools listed above.

Extract a working copy of the legal and operating structure data from the CSBR for review and validation. Use the information observed from the research tools and information sources (listed above) to mark-up the working copy with updates noted in the following two steps.

Legal Structure validation of:

- Percentage of ownership of all partnerships, subsidiaries and related companies.
- Business status of each legal entity in the structure assessing for possible dissolutions, legal name changes and amalgamations.
- Main industrial activity and operating address for those entities representing both the legal and the operating entity of the unit.
- Legal acquisitions or restructuring.
- Preparation of consolidated statements and the participants of the consolidation.
- Foreign ownership.

Note: Administrative data sources are used for automated processing of monthly, quarterly and annual tax data on all legal entities. Accordingly, manual updates are not processed for financial data fields on the legal entities of a structure. However, analysing changing reporting trends can help to indicate business increases or decreases and legal or operational restructuring.

Operating Structure validation of:

- New locations
- Closure of locations
- Sale of operations
- Operational restructuring

Operating Entity validation of:

- ‣ Operating name
- ‣ Operating address
- ‣ Industrial activity
- ‣ Reporting capability for each Responsibility Centre
- ‣ Turnover*
- ‣ Employment numbers*

*Size measure allocation to all operating entities will be finalized during the data coherency phase. The Profiler should, however, review completed surveys data and administrative data sources for data comparison purposes later in the profiling process.

Data assessment phase

The Profiler reviews the working copy, which includes the planned updates, to assess if anything is missing to complete the profile.

Programme Impact Assessment:

The Profiler must assess the impact that the statistical structure changes will have on economic statistics programmes and notify the affected programme areas accordingly. Programme areas are given an opportunity to review and question the proposed changes.

Respondent Contact Assessment:

Contact with respondents should only be made when all research has been completed and it has been determined that insufficient information has been gathered to make all the appropriate changes on the CSBR.

The general rule is to limit respondent contact to no more than once a year whenever possible.

Preparing for Respondent Contact (as required):

Initial contact with respondents is always by telephone. The respondent may request continued contact to be by phone, email, electronic file transfers, registered mail or an on-site visit (in the case of large structures with large impact to economic estimates).

In all contact cases:

- ‣ Prepare a summary document or chart of all structural changes noted. This can be provided to the respondent for updating, when appropriate, or used as a reference guide by the Profiler during contact.
- ‣ Prepare a written list of questions, starting with the most critical information needed to optimize the contact time with the respondent.

Best personnel to contact at a business include Controller (financial, operational), Accountant, Vice-President of Finance, Owner, President or Human Resources.

Other company contacts may be approached upon the request of the initial company respondent.

Data coherence phase

A data coherence analysis is completed prior to applying updates on the CSBR.

Revenue/sales and number of employees:

Vertical coherence – Ensure that reported tax data values at the enterprise level are in line with what is allocated to the operating entities.

Horizontal Coherence – Ensure that the values allocated to the CSBR statistical units are in line with the reported administrative, programme and survey data.

Industrial activity code and description:

If an enterprise has multiple operating entities, ensure that the industrial activity code propagated to the enterprise level appropriately reflects the primary economic activity code of the combined operations.

Report any significant discrepancies in the survey data received to the appropriate programme area for analysis and treatment. Follow-up with respondents may be required to resolve reporting issues or to establish a reporting arrangement to ensure accuracy and coherence of future data.

Data processing phase

Once all information has been reviewed and validated in the steps above, the Profiler begins to process the updates on the live CSBR and completes the profile report.

- ▸ Process all applicable updates.
- ▸ Conduct a final coherence check on all enterprises that have been updated to ensure that changes have been correctly applied on the CSBR.
- ▸ Complete profile report on the CSBR.

Annex F3: Profiling in Colombia

In Colombia, the profiling work of the country's business groups was developed with the objective detailed in section 6.6.4 of these Guidelines. As a result, it has a consolidated base with data collected from public information sources that can be found on web pages such as management reports.

The consolidated information of the business groups located in the country identifies legal units linked to the group holdings and possible SPEs or special purpose vehicles.

The units of the truncated group are also identified and characterized, but the base of affiliated companies is recorded for use when the profiling of the total business group is scaled. The process of identification of the legal units that make up the group in the statistical register of companies or registration of companies in Colombia was carried out with an accuracy rate of 98 per cent. In the work of confirming the results of the profiling, in-person visits were made to two business groups, where the results were presented and discussed.

Annex F4: Profiling in Indonesia

Where activities related to structuring and updating the corporate network are concerned, SBR applications in Indonesia have been developed for several types of users, as follows:

- ▸ Monitoring Supervisor: member of an SBR team that can provide assistance to groups through group supervisors. In the SBR system, monitoring supervisors monitor the progress of the profiling exercise as a whole, in terms of the targets achieved and according to the results of quality control carried out on the profiling for the various subject areas.

> ➤   Group Supervisor: liaison officer responsible for a certain subject area. In the SBR system, group supervisors monitor the progress of group profiling in accordance with the relevant targets, approve the profiling process and check the quality of profiling results from the group members.
>
> ➤   Subject matter area (SMA) operators: profiling officers (profilers) for specific subject areas who can access company directory data across sectors of economic activity. Their activities are under the supervision of group supervisors and monitoring supervisors.

The monitoring structure for the profiling exercise is shown in figure F4.1.

Figure F4.1
Monitoring structure for profiling
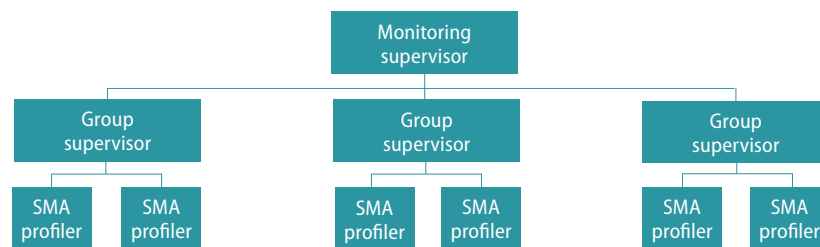
The process of updating and establishing the network structure for profiling companies passes through the following stages:

Determine the group to be profiled. Currently BPS uses the Globe Asia Magazine, which publishes the top 100 enterprise groups in Indonesia based on revenue.

Download the latest annual report on the company's official website, generally to be found in the Investor Information / Company Profile / Business Line / Our Contact / Site Map section.

> ➤   If annual reports are not available, use financial statements.
>
> ➤   If annual reports / financial statements are not available, use the information available on the company's website, generally to be found in the Company Profile / Subsidiary / Business Line menu.
>
> ➤   If there is no official information available, use other Internet information sources, such as YellowPages, ministry websites, Indonesia Stock Exchange websites, etc.

1.   Classify statistical units.
2.   Establish the structure of the company.

> ➤   Prioritize subsidiaries
>
> ➤   Mark duplicate units

Repair / update data (editing / updating)

> ➤   Collect / update contact persons
>
> ➤   Data tidiness (data cleaning): variable entries, writing procedures, completing blank fields
>
> ➤   Checking directory coverage

Obtain approval (specifically group supervisors)

> ➤   New company approval
>
> ➤   Approval of relations

> ‣ Approval for removing relationships
> ‣ Duplicate agreement
> ‣ Approval of editing

Annex F5: Profiling in the Kingdom of the Netherlands

The profiling method, concepts and organization used in Statistics Netherlands (CBS) are described below for both the national and the international part of the businesses.

Organization of profiling

Profiling provides a means of analysing and describing the legal, operational, financial and statistical structure of enterprise groups and of enterprises in order to have a good basis for business statistics.

Because the largest companies in a country have the most influence on the figures in statistics it is wise to concentrate efforts on compiling good-quality profiles for the largest enterprise groups.

In 2005, CBS defined a population of the largest most influential and complex enterprise groups in the Kingdom of the Netherlands, which are called the Top-X groups. The X=1900 enterprise groups were the best at that time. These 1900 also include enterprise groups consisting of only one enterprise (which, by definition, are officially "no enterprise" groups).

Past studies demonstrate that the choice of these groups is not conducive to good quality when an algorithm creates and updates these units.

The Top-1900 enterprise groups were then split in two subsets. One subset includes the so-called ConGO groups, from "Consistentie Grote Ondernemingen" ("Consistency Large Enterprise") groups. These are the 300 groups which are the most important and for these units consistency checks are carried out. The other subset comprises all other units that are profiled but for which consistency checks are not carried out.

All other enterprise groups outside the Top-1900 are not regularly profiled. The business register algorithm derives the structure of these groups. Only at the end of each month, when a new statistical frame is produced by a macro analysis, are units found to have significant changes and these changes are checked manually by profilers.

The profilers for the ConGO groups are placed in the ConGO unit, because these profilers should have good communication lines with the data analysts and account managers. At present there are four profilers in the ConGO unit.

The profilers for the non-ConGO groups (i.e. other Tocongop-1900 groups and reactivated non-Top-1900 groups) form part of the business register department. At present, this comprises eight profilers for the other Top-1900 groups and four profilers for non-Top-1900 groups, but all; of these are employed on a full-time basis.

Figure F5.1
Breakdown of the business register
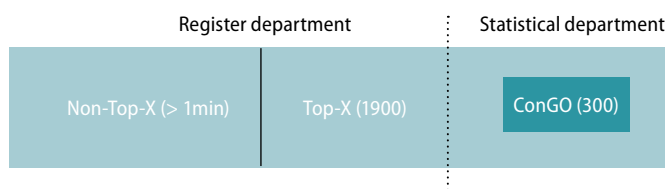
Profiling process in the ConGO unit

This process does not differ significantly from the profiling for the other Top-X groups. The two main differences are:

➤ There are more changes to the larger most complex groups in the ConGO unit than for the other Top-X groups.

➤ The profilers in the ConGO unit have better contacts with colleagues from statistics departments than do profilers for the other Top-X groups. The profilers in the ConGO unit work in a team with a range of expertise. The profilers for the other Top-X groups work in a team consisting only of profilers and NACE code specialists.

The profiling process in the ConGO unit is described first at the national level, and then the international level.

National part of the profiling process

Population

Although the profiling process is a continuous process, some parts of the process are yearly subprocesses. For example, the definition of the population under profiling is revised every year. At CBS there is an algorithm that calculates a score from 0 to 10 for each enterprise group in the business register. The algorithm is known as the complexity and statistical impact – or CSI-algorithm – and it takes into account:

➤ The complexity of the group (such as the number of legal units, control relationships, layers in the cluster of control, number of fiscal units).

➤ The size of the group (grouped into different size classes by either the number of employees or by the balance sheet total).

➤ The importance of the group for statistics (namely, how much the group and its enterprises contribute to its NACE code).

The algorithm is used as the basis for defining the Top-X and ConGO population. Since an algorithm cannot accommodate every criterion, the register department, the ConGO unit and the statistical users can propose individual enterprise groups to be included in the Top-X and ConGO population. For example, for the ConGO population an additional criterion for the inclusion of groups in the population is the presence of international relations with a foreign parent and/or foreign subsidiaries, because globalization issues are an important source of inconsistencies. In addition, an unsatisfactory relationship between the group and CBS can be a reason to include the group in the ConGO population. Additional work can be carried out to improve the quality of statistical figures.

Efforts are made to ensure continuity in the population but there is also need for it to be refreshed.

Sources for profiling

The sources that are being used for profiling are various. The data that they use come with different time stamps and it is the task of the profiler to create the best possible current situation of the group and process this in the business register. The sources that are used include the following:

➤ Trade register information (for existence of and information about legal units and their sole ownership).

- ‣ Tax register (for the relation between fiscal units and legal units, for information about turnover and wages).

- ‣ Annual reports (consolidated or non-consolidated when available).

- ‣ Information from statistical departments (from surveys indicating that the delineation or classification of the group might be wrong or outdated).

- ‣ There are people in the ConGO unit with an account management role. These people are responsible for consistency in statistical data. Inconsistencies can be caused by a wrong structure. In addition, news or future developments will be communicated by the account manager to the profiler.

- ‣ Statistical analysts in the ConGO unit that analyse the data from surveys. Irregularities spotted in this process are communicated to profilers if they are related to the structure of a group.

- ‣ Internet, press releases on the websites of the company itself or on financial and business sites.

- ‣ The AFM[86] register (for ownership and control relations for listed companies).

[86] The AFM promotes the conscientious provision of financial services to consumers and supervises the honest and efficient operation of the capital markets.

Types of profiling

There are two types of profiling: active and reactive.

The reactive process is a continuous process. Businesses are constantly changing, as too are their data and group structure. If there is a change in any of the above-mentioned sources, a profiler must investigate the correctness of the information and ascertain the consequences for the group structure and statistics.

The information received from the trade register (daily) and tax register (monthly) is processed automatically in the business register. There are some safety rules that prevent certain data derived from profiling from being overwritten. The profiler must check if the data have been processed correctly by the batch system. The other sources are not automatically processed.

A statistical frame is created monthly in the business register. All data that have been approved by the profilers are recorded in the new frame. The reactive process can thus lead to 12 different monthly views on a group. For short-term statistics this is essential.

The active process is a time-consuming process that can take several days for one group. This profiling process is explained in the next section.

Active profiling

As indicated in the definition of profiling, a profiler's task is to create the best possible structure of a group.

The first step is to create the cluster of control to delineate the (national) enterprise group. Because the Chambers of Commerce send the legal units from their trade register to CBS on a daily basis, there is no need for the profilers to verify if all legal units are in the business register. Unfortunately, however, the trade register has only recorded sole ownership relations. In many cases, these lead to control relations. If not, a profiler needs to investigate further. In addition, non-sole ownership (i.e. >50% and <100per cent) is not recorded in the trade register. The information in the tax register can also be of additional use. To be able to form tax units a company has to meet certain criteria, one of which is control. The fact that more legal units form a tax unit is therefore additional information.

For companies listed on the stock market, information is available about ownership and control at the AFM. If a company exceeds a certain amount of shares or control it is obliged to report this to AFM.

Also useful are the annual reports. Companies are obliged to file a list of group companies with the Chamber of Commerce. These lists sometimes only include the largest subsidiaries and are not complete, however. Added to which, this information is not digital, but comes in the form of a scan of the annual report from which is was taken. Sometimes the list is not available at all.

Another good source is the company itself. Some have organizational schemes including all the legal units. Other companies do not have such schemes or the schemes are outdated.

Sometimes surveys can furnish information about the structure of the group. Although the processing of this information is more akin to reactive profiling, the information is taken into account when making an active profile.

The second step is to check the consolidation cluster. As CBS has a survey on financial data of non-financial enterprise groups, the consolidation cluster should represent the current situation. In its survey CBS asks respondents to provide consolidated data for the structure to be included in the questionnaire.

The consolidation cluster and the control cluster usually coincide, but there are situations where this is not the case. Examples of such situations are where there are foundations, special purpose vehicles or units with a different institutional sector code included in the cluster of control. These may be excluded from the consolidation cluster. On the other hand, joint ventures may partly be consolidated but they are not included in the cluster of control.

The third step is to create the enterprises within the enterprise group. Some ten years ago, the way in which enterprises were formed changed. Previously CBS looked at enterprises bottom-up, meaning that they tried to group the legal units to create enterprises that would consist of more than one legal unit. Currently, CBS looks at them top-down, meaning that it tries to form the enterprises and to link the legal units to it. Although, in theory, the outcome should be identical, the top-down method yields a better statistical structure.

By definition, the enterprises should be autonomous in respect of the production process. The enterprises should also be externally oriented, meaning that their services and/or goods should not be for internal use (i.e. by the enterprise group). The starting point is to check if the enterprise group has made a breakdown itself in segments. The larger companies all have segments. One segment can be too difficult to manage. Some companies have a geographical breakdown; for others the breakdown is by brands; and for others it is by activities. The operational segments have to be investigated, to ascertain: are they autonomous? Are they still autonomous at a lower level (if the segments can be split) or not? Are the products or services they sell outside the (national) group? If not, the segment is not externally oriented and it should be combined with the operational segment(s) for which it provides goods and services. When the combined or split segments are autonomous and externally oriented, the company should be asked if they are able to provide meaningful statistical figures. If they cannot, the segments should be further combined or split.

Once the units are final, the characteristics of these units should be determined (such as NACE code, size class, etc.). This can be done on the basis of the information from the group itself (direct information or information from annual reports) or by aggregating and consolidating information from the legal units (next step).

The fourth step is to link the legal units to the enterprises. If consolidated figures are not available at the enterprise level, a second best is their aggregation from the legal units. Sometimes a complete list of legal units is available for each segment in the annual report. If not, based on the NACE code of the legal units and the enterprises, it should be possible to link all units. Of course, the result can be verified at the company: the company also has the advantage of a good delineation.

The last step is to make a report of the research and the result. The report is essential, because it is the justification of the chosen delineation. It also serves as reference and as a document that is needed when an enterprise group is handed over from one profiler to another. The report is also shared with the other people in ConGO unit (account managers and analysts) and the statistical users.

The last step also includes processing the changes into the business register.

Depending on the changes in a group, a reprofiling exercise could either be a new profile prepared from scratch or an update of the existing profile. CBS tries to profile each enterprise group at least once a year actively. Whether or not this is possible depends on the resources available.

International part of the profiling process

The profiling process for profiling MNEs that are in the EGR dos not differ significantly from the national profiling process. Only the dimension and organization are different. The dimension is different because for international profiling the MNE is the largest unit. The global enterprise group (GEG) consists of one or more global enterprises (GEN) and these consist of one or more truncated enterprises (TEN). The GEG also consists of more than one truncated enterprise group (TEG).

Figure F5.2
International and national units
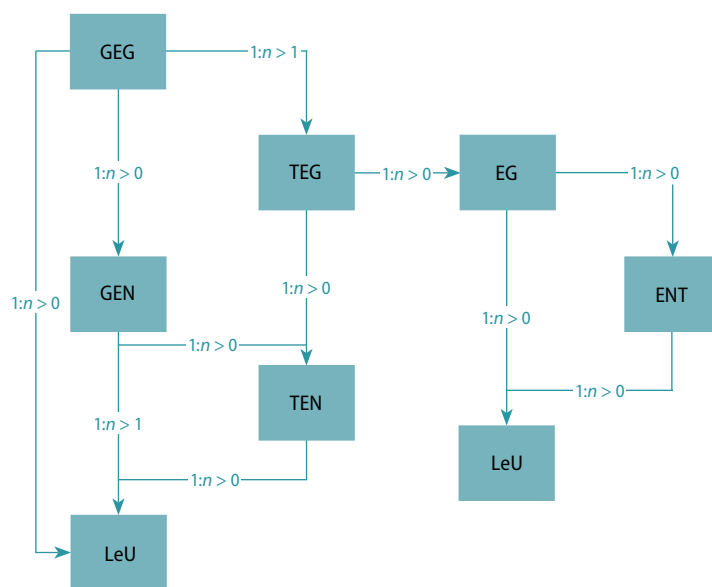
In the ESSnet on profiling, an endeavour has been made to ensure that the national enterprise coincides with the truncated enterprises, but there were numerous cases where this was not possible. It became clear that the TEN is not suitable for national needs to make statistics. Furthermore, the TEG is more a theoretical concept than a useful concept.

In the Kingdom of the Netherlands there are many instances where legal units have the same foreign parent but they all belong to the same global enterprise group. Examples include Scania and Abbott. These are separate enterprise groups in the Kingdom of the Netherlands. If they are combined with a TEG, this would not yield meaningful figures, nor would the group itself be happy to aggregate (and consolidate!) figures. Enterprises are formed within the national enterprise group, not within the TEG.

For combining national with international profiling, therefore, the concepts of TEG and TEN are less useful. The relation between GEG and EG and between global enterprise (GEN) and national enterprise (ENT) is more valuable. When international profiling is adapted as part of the national profiling focusing on the GEG perimeter (cluster of control and consolidation cluster), the focus should be placed on GEN and the relation with the national enterprises.

At the current time, international profiling consists of the following activities:

> For the Top-X enterprise groups the direct foreign parent, the global group head and the ultimate controlling institutional unit/global decision centre are checked yearly. This activity is part of the active profiling.

> Repair work is carried out for the EGR when the data from the national business register are sent to the EGR and combined with other sources. This activity is part of the reactive profiling and not only for Top-X enterprise groups but for all groups in the EGR.

Profiling as part of the ConGO process

Need for a ConGO unit

Since 2010 CBS has had a unit dedicated to delivering consistent microdata on large and complex enterprise groups for a number of business statistics. One of the reasons for the creation of this ConGO unit was the large inconsistencies between statistics that came to light when compiling the national accounts. These inconsistencies were caused by inconsistent microdata for large enterprise groups. It was expected that, by concentrating the work on these enterprise groups in one unit, inconsistencies could be detected and solved early in the statistical process and hence the quality of both individual statistics and national accounts could be improved. After five years of experience, the quality was indeed improved.

With regard to the quality of the statistical results it was considered necessary to reduce the imbalances between the first and final estimates growth rate figures for the Netherlands economy. Over the last 20 years the average adjustment of the growth rate was roughly + 0.5per cent of GDP. Differences between short-term statistics and the annual production statistics were one of the causes for these adjustments.

Simply because of their size the enterprise groups in the ConGO population make an important contribution to the Netherlands economy: the 300 largest non-financial enterprise groups (having more than 2,000 enterprises and more than 10,000 legal units) account for 50 per cent of the total value added, 55 per cent of the balance sheet total and 40 per cent of the turnover of all non-financial enterprise groups in the Kingdom of the Netherlands.

This means that incorrect or inconsistent data for these enterprise groups have a major impact on the outcome of individual statistics or even on the national accounts. Unfortunately, until 2010, most inconsistencies between individual statistics showed up only at the compilation stage of the national accounts. At that moment the related statistics had often already been published by the different departments of CBS, which tended to operate in isolation. Consequently, suggestions by national accounts compilers to correct these inconsistencies came at such a late stage that they could no longer be processed in the individual statistics. Moreover, at that point, the reference period in respect of which questions had arisen about the data was too far back in time for the questions reasonably to be put to the enterprise groups. As part of the redesign programme, it was considered whether it might be feasible to solve inconsistencies at an earlier stage of the statistical process or even to prevent their occurrence. One means of achieving this goal could be to concentrate all the work on the large enterprise groups within one organizational unit that would be responsible for delivering consistent data on those enterprise groups and their enterprises to the statistical departments responsible for the compilation of the different statistics.

ConGO consistency tool

In cooperation with the National Accounts Department a number of rules were formulated with which data from different statistics had to comply. In essence these rules can be considered as the translation to the micro level of the confrontations carried out by national accounts at the macro level. All these rules were implemented in a software application called the "ConGO consistency tool".

This tool is fed daily with data from the different statistics involved and indicates at both the level of the enterprise groups and that of the enterprises which consistency rules have been violated. In practice, an expected equality between two values is translated into a percentage with which the relative difference has to comply. For a chosen reference period the software supplies a consistency matrix at enterprise group level that shows the value of variables observed at that level, both for the sum of the enterprises and for the individual enterprises. This matrix usually serves to give the account manager a first impression of the degree of consistency of the enterprise group and an indication of the possible sources of inconsistencies.

Consistency is not only checked for a fixed reference period, but also in time, such as the consistency between data for the same enterprise from short-term (monthly, quarterly) statistics and from annual statistics. In addition, the longitudinal development of indicator ratios is considered as an aid in detecting deviating behaviour by an enterprise.

The following statistics are included in the consistency tool: finances of enterprise groups, SBS, STS, industrial production statistics, investment statistics, international trade in goods and services, research and development, and statistics on employment and wages. The data on corporate tax and VAT are also included.

The variables for which consistency rules are defined are the following: depreciations, earnings, employee benefit costs, wages, export, import, intermediate consumption, investments, net financial result, net sales and other operating revenues, number of employees, production, research and development expenses and turnover.[87]

Organization of the ConGO unit

The creation of the ConGO unit is the result of a long-drawn-out process that started in the late 1990s when the role of account manager was created within the field service of CBS. At that time these account managers were responsible for making arrangements with enterprise groups on the delineation of enterprises (the statistical unit for

[87] For more information, see Vennix, K., ICES IV, 2012, The treatment of large enterprise groups within Statistics Netherlands, 2012).

most surveys of CBS) within the groups and the manner in which to approach these enterprises for data collection. This means that their work was predominantly focused on the profiling of enterprise groups and not on the statistical data supplied by the enterprise groups and their enterprises. Over the past 15 years the role of account manager has gradually shifted its focus from profiling work to statistical work. Concurrently the function of account manager has become a role in the statistical departments of CBS, rather than in the field service. Today the main task of the account manager is to deliver consistent data for a number of statistical areas for the enterprise groups in his portfolio.

The unit consists of five account managers, four profilers, eight statistical analysts on enterprise level statistics, six statistical analysts on enterprise group level and five staff members performing assisting, technical or managing roles.

The unit is now responsible for maintaining relations with the enterprise group, profiling the structures in the business register, processing microdata and delivering consistent microdata to the statistical departments. A good knowledge of the enterprise group and good relations with it are of vital importance in this context. Accordingly, account managers are expected to visit their enterprise groups at least once every year. If necessary the profiler will visit the enterprise group together with the account manager.

Task of the profiler in the ConGO unit

The role of the profiler would be to update the business register and it would be logical that the profiling team is part of the business register staff. For the non- ConGO units, this is the case. But for the ConGO unit, there is a need for close cooperation with the account manager(s) and the statistical analysts.

If profilers spot a major change in a group, they consult the account manager regarding how to process this in the business register. This decision has an impact on the account manager's workload later on. Sometimes this is difficult, because the account manager's responsibility is to ensure data consistency for a previous period (T-1 or even T-2) and the profilers' main concern is the actual structure of the group in period T.

Statistical analysts are the first people that could detect wrong figures in statistics. As things stand, they are focused on one statistic at a time, but they could also check statistics in relation to other statistics. That would simplify matters for the account managers (i.e. leaving fewer inconsistencies for them to explain or correct) and inconsistencies would be picked up at an even earlier stage. At the moment the statistical analysts are doing checks for each statistic by checking, for example, the current value T with the value of T-1. When there is a change that they do not expect, they consult the profiling reports. If this gives no answer for the unexpected change they contact the group for an explanation. If the explanation is that there was an incorrect delineation of the statistical units in the group, the statistical analyst will inform the profiler, who will correct it in order that the next time a questionnaire is sent to the group the correct structure is used.

The contact between the account manager and the profiler is similar to that between the statistical analyst and the profiler: if account managers detect inconsistencies caused by a wrong structure they contact the profiler to correct the structure in the register. In addition, the account manager may request an improved delineation of the enterprises in the group. This may arise from a request sent to the account manager by the group itself or from an analysis of the data of the group. Furthermore, a more efficient data collection may be a reason to request a change in the delineation of the enterprises in the group.

The role of the profiler in the ConGO process is therefore twofold: first, to create the best possible delineation of the enterprise group and its enterprises before questionnaires are sent; and, second, to correct errors in the delineation handed to them by the statistical analysts and account managers

Figure F5.3
Communication lines
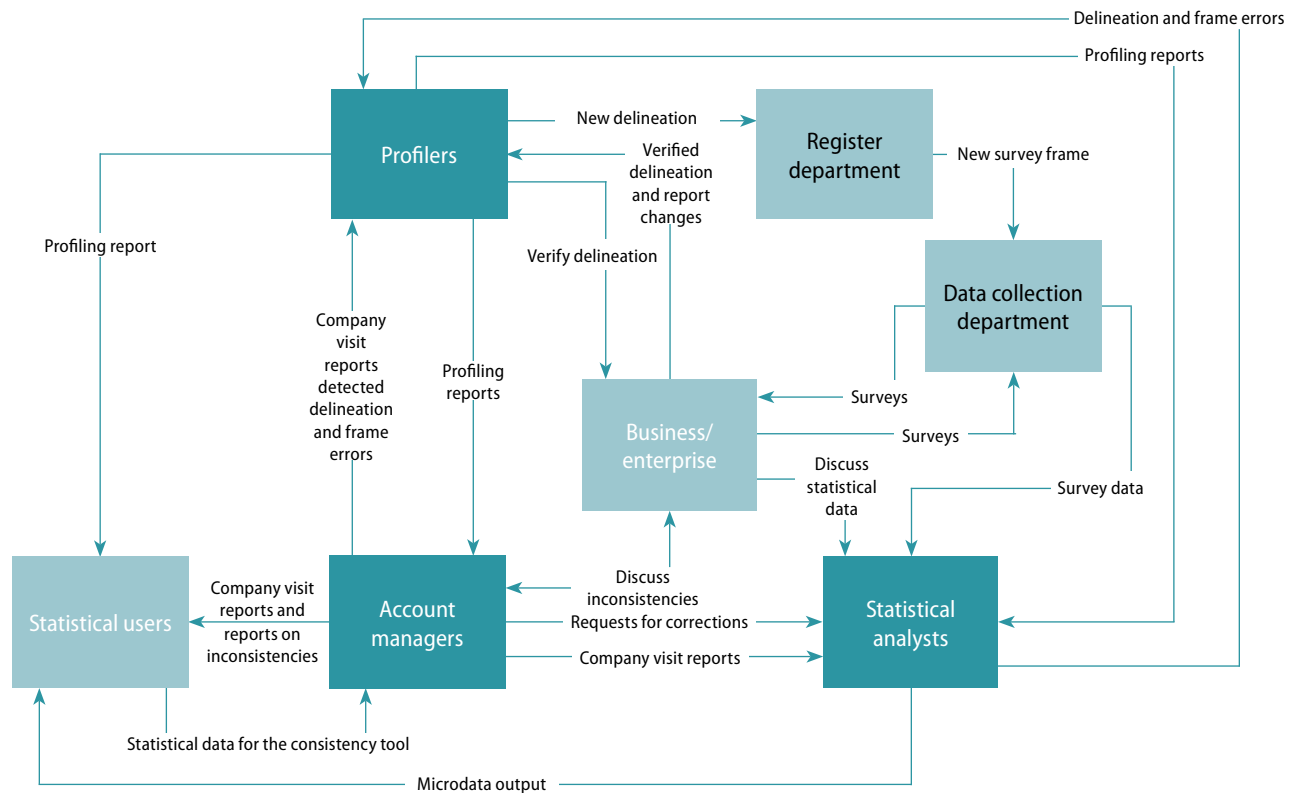
Future plans, improvements

The ConGO unit has now been in existence for more than five years, but a continuous learning and improving process is still under way, and plans are in place to further test and implement the system. Below is the description of some of these future plans.

One person for one group

CBS would like to see the statistical analysts carrying out their investigation not only statistic by statistic but also performing some consistency work by looking at more statistics at the same time. This is a prelude to a situation where all three different roles are combined in one person: one and the same person could carry out the profiling, the statistical analysis and also the consistency work. In this way, that person will know about everything relating to the group. This has an advantage in that it will speed up the process of making an enterprise group consistent, but it also has a disadvantage in that knowledge will be concentrated in one person and if this person is unavailable for some reason it will be difficult for colleagues to take over. This approach will be tested first before deciding that it is a better way of ensuring consistency in statistics.

Splitting the Top-X

The Top-X units consist of a mix of complex units and units of importance for statistics. A proposal is being studied to split the Top-X and create two populations, each for its own needs: a Top-P for enterprise groups that need profiling and a Top-S for groups that are most important for statistics. Many enterprise groups in the current Top-X are there because they are important for statistics but they are easy units (Enterprise group = Enterprise) and stable units. These units (some 700 in all) can be put in the Top-S, making the Top-P smaller (around 1,200). The Top-P will then be a subset of Top-S because it makes no sense to profile units that are statistically insignificant. This new view will give both the profiling department and the statistical department more freedom to add and remove units from the Top population.

ConGO-light

For the ConGO population of about 300 enterprise groups inconsistencies are solved or explained. It is unrealistic to assume that there are no (large) inconsistencies in other units. A second consistency tool could be developed and fed with all the data from the Top-X (or Top-S, if it is decided to split the Top-X) enterprise groups. This can be used as a monitoring system. For groups with many or large inconsistencies, it may be decided to add those units to the ConGO population. Likewise, important units may be removed from the ConGO population if they are consistent each year and placed in the Top-X outside the ConGO population. The second consistency tool, used to monitor this unit, would remain in use.

State-of-the-art profiling method

Describing a state-of-the-art profiling method is difficult. The needs for quality vary between statistical institutes. There might be a joint programme under which statistics have to be published, but the national needs and hence its quality may differ. The available resources will also vary from institute to institute. Instead of describing a state-of-the-art profiling method it is better to describe the key success factors for good profiling. The focus areas of the profiling and the way in which it is organized may be filled in nationally.

The key success factors for good profiling are:

- Define a population to profile actively. Do not profile each unit actively.
- Make sure that the most important groups for national and international statistics are included in this population.
- For groups that are not to be profiled actively, create a procedure whereby significant changes are only made when batch processing these units and these will be checked by profilers.
- Update the profile regularly. Frequency may vary depending on the kind of units in the population.
- Profile at least at the level of the national enterprise group, and preferably that of the global enterprise group
- Make sure that the delineation in units within a group is both suitable for statistics and for the company.
- Make sure to establish a good contact with the group.
- Make sure that the result of the profile is implemented in both the national register and (indirectly) in the EGR.

> ▸　Communicate with the statistical users. Try to predict the consequences of a new (drastically changed) delineation for the statistics and. when not sure, contact them to discuss the proposal.

> ▸　During profiling use as much information available as possible, creating the best view of the group before contacting the group.

> ▸　During profiling work top-down; this leads to a better delineation of the enterprises than the bottom-up approach. Annual reports and the company's website often give information about how the company is organized (segments).

> ▸　Stay in contact with the FATS team when a change of the ultimate controlling institutional unit is detected.

> ▸　Send profiling reports to all stakeholders

> ▸　Process corrections from statistical departments or the company as swiftly as possible but in a coordinated manner.

Annex F6: Profiling in South Africa

The example of South Africa refers mainly to the working steps and documentation of the profiling process. The example shows a part of the documentation spreadsheets that have to be filled in by the profiler for each enterprise group. Figure F6.1 shows a summary list of the various documents which have to be filled-in.

Figure F6.1
Summary list of documents

Name of EG
EN Number
Name of Profiler
Name of Supervisor

	Correctly placed in file		Completed in full	
Order of the EG file	**Yes**	**No**	**Yes**	**No**
Profiling Summary Sheet				
Annual report and/or other financial information received from respondent				
Organogram/ Organizational Structure				
Profile Spreadsheet				
Classification Approval Cover Page				
Quality self-assessment checklist completed by the Business Register Analyst				
All communication between the Business Register Analyst and the respondent:				
√ Profiling Action Sheet				
√ Request for profiling form				
√ Emails (Survey areas and companies)				
√ Faxes (if applicable)				

√ Interview agenda
√ Interview Outcome Report
√ Data request form (if applicable)
Checked by:

Figures F6.2 to F6.5 refer to the quality self-assessment checklist completed by the business register analyst. This checklist is fairly comprehensive and shows the single steps that are to be considered in the profiling process in great detail.

Figure F6.2
Profile quality self-assessment checklist – cover page

Name of EG
EN Number
Name of Profiler
Name of Supervisor

The Profile Quality Checklist uses a process approach in the sense that the production of a profile is viewed as a sequence of processes that result in the final product, namely the profile. The five basic processes are defined as:

> ➤ Initiation of profile
> ➤ Analysis of current situation
> ➤ Arrangement, conduct and follow-up of profiling interview
> ➤ Development of revised profile
> ➤ Dissemination and use of revised profile.

This particular choice of processes is somewhat arbitrary in the sense that the individual profiling activities of which they are composed could be combined in a variety of different ways. The most important consideration is that all key activities are included.

The checklist is in three parts.

Part 1 Comprises general information about the enterprise group that is the subject of the profile and about the management of the profile.

Part 2 Comprises a set of questions about the process that created the particular profile. The answers contribute not only to an assessment of quality and performance of the profiling operation but also to an assessment quality of the final product, the profile.

Part 3 Comprises an assessment of the profile in terms of the output quality characteristics – relevance, accuracy, timeliness, accessibility, interpretability and coherence – defined in the South African Quality Assessment Framework. The remaining SASQAF characteristics – prerequisites of quality and methodological soundness, and integrity – are assessed through the process analysis in Part 2.[88]

Key terms and abbreviations are defined in the Profiling Glossary, which is currently in development. They include the following:

Legal unit (LU)	Enterprise Group (EG)
Operating Unit (OPU)	Enterprise (EN)
Income Tax Unit (IT)	Kind-of-Activity Unit (KAU)

[88] South African Statistical Quality Assessment Framework (SASQAF), second edition. Available at www.statssa.gov.za/standardization/SASQAF_Edition_2.pdf.

Value Added Tax Unit (VAT)	Geographical Unit (GEO)
Pay As You Earn Unit (PAYE)	Observational Unit (OU)

Figure F6.3
Profile Quality Self-assessment Checklist, part 1 – background information about the enterprise group

Name of Enterprise Group				
Is holding company listed: ☐ Yes ☐ No		Financial year end:		
Size and complexity of EG (numbers when profile initiated)				
Group Enterprise Turnover:		Group enterprise No. of employees:		
No. of LU registered in South Africa		No. of ITs:		
No. of active LU registered in South Africa		No of VATs:		
No. of PAYEs:		No. of ENs:		
No. of KAUs:		No. of GEOs:		
Is the holding company at head of EG a pure investment holding company ☐ Yes ☐ No				
Surveys for which EG in scope (when profile initiated)		Year	Survey Code	No. of OUs
Annual financial survey	☐ Yes ☐ No			
Quarterly financial survey	☐ Yes ☐ No			
Quarterly employment survey	☐ Yes ☐ No			
Manufacturing production	☐ Yes ☐ No			
Wholesale	☐ Yes ☐ No			
Monthly retail survey	☐ Yes ☐ No			
Monthly motor trade	☐ Yes ☐ No			
Monthly accommodation	☐ Yes ☐ No			
Monthly food and beverage	☐ Yes ☐ No			
LSS – Construction	☐ Yes ☐ No			
LSS – Mining	☐ Yes ☐ No			
LSS – Agriculture	☐ Yes ☐ No			
LSS – Transport	☐ Yes ☐ No			
LSS – Electricity	☐ Yes ☐ No			
Other (specify)	☐ Yes ☐ No			
Other (specify)	☐ Yes ☐ No			

General Information about Management of the Profile

Name and position of lead profiler
Names of other LBU staff helping with profile
Date profiling initiated
Date profile completed
Staff days spent by lead profiler
Staff days spent by other LBU staff
Staff days spent by other staff
Any general comments about EG or profile

Figure F6.4
Profile quality self-assessment checklist, part 2 – process perspective

A.   Initiation of profile

Was this a new profile or an update of an existing profile?

☐ New profile   ☐   Existing profile

What was (were) the reason(s) for the profile?

(a) Request from one or more survey areas?	☐ Yes ☐ No
If yes, which area(s)?	
(b) Triggered by information from administrative source(s)?	☐ Yes ☐ No
If yes, which sources(s)?	
(c) On schedule provided by LBU Management?	☐ Yes ☐ No
(d) Other reason(s)?	☐ Yes ☐ No
If yes, please specify	

Were survey areas consulted regarding the necessity for profile?

☐ Yes ☐ No

If yes, which area(s), and what was (were) their viewpoints?

B.   Analysis of current situation

What information sources were used in analysing current situation and preparing for the interview?

(a) Documentation from previous profile?	☐ Yes ☐ No
(b) Statistical Business Register (SBR)?	☐ Yes ☐ No
If yes, what information was used?	
(c) Administrative Business Register?	☐ Yes ☐ No
If yes, what information was used?	
(d) Commercial Sources (e.g. MacGregors, Top 200, Who Owns Whom)?	☐ Yes ☐ No
If yes, which sources and what information was used?	
(e) Annual report for Enterprise Group?	☐ Yes ☐ No
If yes, what information was used?	
(f) Survey areas?	☐ Yes ☐ No
If yes, which survey area(s) and what information did they provide (e.g. lists of observational units, reporting problems, questionnaires)?	
(g) Other source(s)?	☐ Yes ☐ No
If yes, what sources(s) and what information did they provide?	

How was analysis and technical preparation for interview undertaken?

(h) Describe analysis process	
(i) Were survey staff involved?	☐ Yes ☐ No
If yes, who, and what was (were) their contribution(s)?	
(j) Was information from previous profile consistent with information in SBR?	

What documentation was prepared before interview?

(k) Organigramme	☐ Yes ☐ No
If yes, did it contain LUs, OPUs, a mixture of both, or something else?	
(l) List of LUs?	☐ Yes ☐ No
If yes, how was it obtained?	
(m) IT numbers?	☐ Yes ☐ No
If yes, were their relationships to LUs indicated, and if so how?	
(n) List of VAT numbers?	☐ Yes ☐ No
If yes, were their relationships to LUs indicated, and if so how?	
(o) List of PAYE numbers?	☐ Yes ☐ No
If yes, were their relationships to LUs indicated, and if so how?	
(p) List of OPUs?	☐ Yes ☐ No
If yes, were their relationships to LUs indicated, and if so how?	
(q) List of ENs?	☐ Yes ☐ No
If yes, were their relationships to LUs and/or OPUs indicated and if so how?	
(r) List of KAUs?	☐ Yes ☐ No
(s) List of GEOs?	☐ Yes ☐ No
If yes, were their relationships to LUs and/or OPUs indicated and if so how?	
(t) Lists of OUs for relevant surveys?	☐ Yes ☐ No
If yes, for which surveys?	
(u) Were all OUs verified with the relevant survey areas?	☐ Yes ☐ No
Copies of questionnaires for relevant surveys	☐ Yes ☐ No
(v) Problems encountered by relevant surveys	☐ Yes ☐ No
Profile spreadsheet?	☐ Yes ☐ No
(w) List of questions for interview?	☐ Yes ☐ No
(x) Other documentation?	☐ Yes ☐ No
If yes, what did it contain?	

C. Initial profiling interview arrangements, conduct and follow-up

If more than one interview took place during the course of the profiling operation, please complete Section D for each interview.

What were arrangements for the initial profiling interview?

(a) Were the objectives of the interview explicitly documented?	☐ Yes ☐ No
If yes, in what form?	
(b) What was the position of the contact person in the Enterprise Group?	
(c) Was contact person already known from previous profiling interview?	☐ Yes ☐ No
If no, how was contact person identified?	
(d) Was documentation sent in advance to contact person?	☐ Yes ☐ No
If yes, what documentation?	
(e) Was the length of interview agreed in advance?	☐ Yes ☐ No
If yes, what was the agreed length?	Minutes
If yes, was the agreed length less than originally requested?	☐ Yes ☐ No

How was the initial profile interview conducted?

(f) How many EG staff participated in interview?	

(g) What was (were) their position(s)?	
(h) How many LBU staff participated in interview?	
(i) Did any other Stats SA staff participate in the interview?	☐ Yes ☐ No
If yes, who, and what was their role?	
(j) How long did the interview last?	Minutes
(k) Would a longer period have been useful?	☐ Yes ☐ No
If yes, how much longer and why?	Minutes
(l) Did interview determine if there had been any major organizational changes – merger, split offs, break up, etc.	☐ Yes ☐ No
(m) Did interview involve discussion of LUs?	☐ Yes ☐ No
(n) Did interview include request for list of dormant LUs	☐ Yes ☐ No
(o) Did interview involve discussion of OPUs?	☐ Yes ☐ No
(p) Did interview involve discussion of IT units?	☐ Yes ☐ No
(q) Did interview involve discussion of VAT units?	☐ Yes ☐ No
(r) Did interview involve discussion of PAYE units?	☐ Yes ☐ No
(s) Did interview involve discussion of economic activities descriptions?	☐ Yes ☐ No
(t) Did the interview involve discussion of any surveys?	☐ Yes ☐ No
If yes, which surveys and what was discussed?	
(u) Did the interview involve discussion of OUs?	☐ Yes ☐ No
(v) Were the interview discussions and conclusions documented?	☐ Yes ☐ No
(w) Were all the objectives of the interview achieved?	☐ Yes ☐ No
If no, what was not achieved and why?	

What were interview follow-up arrangements?

(x) Did EG agree to provide additional information after interview?	☐ Yes ☐ No
If yes, what information and was it received?	
(y) What other follow-up arrangements (if any) were made with EG?	
(z) Was an interview report prepared?	☐ Yes ☐ No
If yes, was interview report sent to the Enterprise Group for verification and/or comments?	☐ Yes ☐ No
If yes, was a response received and if so what was its content?	☐ Yes ☐ No
(aa) Was any other documentation sent to EG for information, verification and/or comments?	☐ Yes ☐ No
If yes, what documentation and was a response received?	
(bb) Was interview report prepared and sent to staff in affected survey areas?	☐ Yes ☐ No
If yes, were any responses received, and if so, what were they?	☐ Yes ☐ No
(cc) How many follow-up interviews were conducted after initial profile interview?	

D. Development of updated profile

Note 1: Changes should be relative the previous profile spreadsheet, if there was one, otherwise everything is a change)

How were results of profiling and follow-up analysed?

(a) Describe analysis process	
(b) Were survey staff involved in analysis?	☐ Yes ☐ No
If yes, what role did they play?	

(c) Were procedures in Profiling Manual precisely followed?	☐ Yes ☐ No
If no, what were the reasons for the deviations?	

What were the changes to EG profile as recorded in Profile Spreadsheet?
(Note 2: actual changes in SBR covered in later section - F2)

(d) Were there additions/deletions to set of active LUs?		☐ Yes ☐ No
If yes, how many?	additions	deletions
(e) Were there additions/deletions to set of inactive LUs?		☐ Yes ☐ No
If yes, how many?	additions	deletions
(f) Were there additions/deletions to set of ITs?		☐ Yes ☐ No
If yes, how many?	additions	deletions
(g) Were there additions/deletions to set of VATs?		☐ Yes ☐ No
If yes, how many?	additions	deletions
(h) Were there additions/deletions to set of PAYEs?		☐ Yes ☐ No
If yes, how many?	additions	deletions
(i) Were there proposed additions/deletions to set of active ENs?		☐ Yes ☐ No
If yes, how many?	additions	deletions
(j) Were there proposed additions/deletions to set of inactive ENs?		☐ Yes ☐ No
If yes, how many?	additions	deletions
(k) Were there proposed additions/deletions to set of active GEOs?		☐ Yes ☐ No
If yes, how many?	additions	deletions
(l) Were there proposed additions/deletions to set of OUs		☐ Yes ☐ No
If yes, how many?	additions	deletions
If yes, which surveys would be affected and how?		
(m) Were there other proposed changes to set of OUs?		☐ Yes ☐ No
If yes, how many?	classifications	addresses
If yes, which surveys would be affected and how?		
(n) Were there any other changes not mentioned above?		☐ Yes ☐ No
If yes, what were they?		

E.   Dissemination of revised profile information
What documentation was prepared and made available?

(a) New profile summary, indicating main changes with reasons?	☐ Yes ☐ No
If yes, to whom and how was document made available?	
(b) Revised organigramme?	☐ Yes ☐ No
If yes, to whom and how was document made available?	
(c) Revised profile spreadsheet?	☐ Yes ☐ No
If yes, to whom and how was document made available?	
(d) IT numbers with additions and deletions indicated?	☐ Yes ☐ No
If yes, to whom and how were the numbers made available?	
(e) VAT numbers with additions and deletions indicated?	☐ Yes ☐ No
If yes, to whom and how were the numbers made available?	
(f) PAYE numbers with additions and deletions indicated?	☐ Yes ☐ No
If yes, to whom and how were the numbers made available?	
(g) ENs, KAUs, and GEOs with proposed changes indicated?	☐ Yes ☐ No

If yes, to whom and how was document made available?		
(h) Lists of observational units with proposed changes indicated?	☐ Yes	☐ No
If yes, to whom and how was document made available?		
(i) Other documentation?	☐ Yes	☐ No
If yes, what did it contain and how was it made available?		

What changes were actually made to Statistical Business Register (SBR) as result of the profile?

Units associated with EG	Before Profile	After Profile
(j) No. of active ENs		
(k) No. of reactivated ENs		
(l) No. of deactivated ENs		
(m) No. of ceased ENs		
(n) No. of KAUs		
(o) No. of birthed KAUs		
(p) No. of birthed GEOs		
(q) No. of active GEOs		
(r) No. of reactivated GEOs		
(s) No. of deactivated GEOs		
(t) No. of ceased GEOs		
(u) No. of GEOs transferred to other ENs		
Changes in classification		
(v) No. of changes to economic activity classification of active ENs		
(w) No. of changes to economic activity classification of active KAUs		
(x) No. of changes to economic activity classification of active GEOs		
(y) Were the changes in SIC approved by QIU – classification experts?	☐ Yes	☐ No
(z) Is the information on the SBR consistent with the information on the profile spreadsheet	☐ Yes	☐ No

What changes were actually made to SMS as result of profile?

Survey	# OUs before profile	# OUs added activated	# OUs removed deactivate	# OUs after profile	# OUs changed econ classifi-cation	# OUs changed size group	# OUs other change
(aa) Annual Financial Survey							
(bb) Quarterly Financial Survey							
(cc) Quarterly Employment Survey							
(dd) Manufacturing Production							
(ee) Monthly Wholesale							
(ff) Monthly Retail							
(gg) Monthly Motor Trade							
(hh) Monthly Accommodation							
(ii) Monthly Food and Beverage							
(jj) LSS - Construction							
(kk) LSS- Mining							

(ll) LSS - Agriculture	
(mm) Other (specify)	
(nn) Other (specify)	
(oo) Is the information on the SMS consistent with the information on the profile spreadsheet	☐ Yes ☐ No

What changes were actually made to Administrative Business Register as result of profile?

(pp) Number of IT-VAT links changed
(qq) Number of IT-PAYE links changed
(rr) Number of other links changed

What were reactions of survey areas to profile results?

(ss) Were all relevant survey areas made aware of the new profile?	☐ Yes ☐ No
If yes, how?	
(tt) Did any survey area express appreciation for the new profile?	☐ Yes ☐ No
If yes, which survey area(s) and what benefits did they feel they had derived?	
(uu) Did any survey area express dissatisfaction with new profile?	☐ Yes ☐ No
If yes, which survey area(s) and what problems did they feel they had encountered?	

Figure F6.5
Profile quality self-assessment checklist, part 3 – product perspective

F.   Relevance

(a) Was profile initiated in response to survey area request?	☐ Yes ☐ No
(b) Was all relevant information reviewed prior to interview?	☐ Yes ☐ No
(c) Was initial list of LUs prepared?	☐ Yes ☐ No
(d) Was initial list of ITs prepared?	☐ Yes ☐ No
(e) Was initial list of VATs prepared?	☐ Yes ☐ No
(f) Was initial list of PAYEs prepared?	☐ Yes ☐ No
(g) Was initial list of ENs prepared?	☐ Yes ☐ No
(h) Was initial list of KAUs prepared?	☐ Yes ☐ No
(i) Was initial list of GEOs prepared?	☐ Yes ☐ No
(j) Was initial list of OUs prepared?	☐ Yes ☐ No
(k) Was initial information verified with survey areas?	☐ Yes ☐ No
(l) Were survey staff involved in profiling interview preparation?	☐ Yes ☐ No
(m) Were survey staff involved in profiling interview?	☐ Yes ☐ No
(n) Were survey staff involved in analysis of results?	☐ Yes ☐ No
(o) Did profile result in changes in number of active ENs?	☐ Yes ☐ No
(p) Did profile result in changes to economic active classifications of active ENs?	☐ Yes ☐ No
(q) Did profile result in changes in number of active GEOs?	☐ Yes ☐ No
(r) Did profile result in changes to economic active classifications of active GEOs?	☐ Yes ☐ No
(s) Did profile result in substantial changes to OUs?	☐ Yes ☐ No
If yes, which surveys and what changes?	
(t) Did supervisors express satisfaction with profile?	☐ Yes ☐ No
(u) Did survey managers express satisfaction with profile?	☐ Yes ☐ No

| If yes, which surveys and how? | | |
| (v) What other evidence is there that the profile was relevant? | | |

G. Accuracy

(a) Were major changes to profile verified by EG contact person?	☐ Yes	☐ No
(b) Were major changes to profile verified by supervisor?	☐ Yes	☐ No
(c) Were major changes agreed with affected survey areas?	☐ Yes	☐ No
If yes, which areas?		
(d) Was list of LUs verified?	☐ Yes	☐ No
If yes, how?		
(e) Was list of ITs verified?	☐ Yes	☐ No
If yes, how?		
(f) Was list of VATs verified?	☐ Yes	☐ No
If yes, how?		
(g) Was list of PAYEs verified?	☐ Yes	☐ No
If yes, how?		
(h) Were ENs derived in accordance with Profiling Manual?	☐ Yes	☐ No
(i) Were GEOs derived in accordance with Profiling Manual?	☐ Yes	☐ No
(j) Were lists of OUs verified with survey areas?	☐ Yes	☐ No

H. Timeliness

(a) What was the length of time between profile initiation and completion?	days
(b) What was the length of time between initial profiling interview and profile completion?	days
(c) How long did it take to prepare for the interview documents	days
(d) Was a target date for completion of profile set at the time profile initiated?	☐ Yes ☐ No
If yes, was profile completed by this initial target date?	☐ Yes ☐ No
(e) What other evidence is there that the profile was timely?	

I. Accessibility

(a) Were staff in affected survey areas made aware of conduct of profiling operation from initiation to completion?	☐ Yes	☐ No
If yes, what is the evidence?		
(b) Have profile results been made readily accessible to survey areas?	☐ Yes	☐ No
(c) Were survey staff in affected survey areas made aware of new profile summary?	☐ Yes	☐ No
If yes, what is the evidence?		
(d) Have any survey area staff accessed new profile summary?	☐ Yes	☐ No
If yes, what is the evidence?		
(e) What other evidence is there that the profile was accessible?		

J.   Interpretability

(a) Did survey area staff in affected areas understand reasons for conduct of profiling operation?	☐ Yes ☐ No
If yes what is the evidence?	
(b) Do survey area staff in affected areas understand new profile?	☐ Yes ☐ No
If yes what is the evidence?	
(c) Does profile summary explain the main changes in profile?	☐ Yes ☐ No
(d) What other evidence is there that the profile is understandable?	

K.   Coherence

(a) Are changes from previous profile explicable?	☐ Yes ☐ No
If no, what is the problem?	
(b) Is revised profile consistent with survey OUs in affected survey areas?	☐ Yes ☐ No
If no, how are inconsistencies explained?	
(c) What other evidence is there that the profile is coherent?	

Annex F7: Profiling in Switzerland

Working on the one hand with the country's large companies and administrations and on the other hand with the internal partners of the FSO, profiling acts as a bridge between the economic and statistical world. Faced with two constantly evolving realities, it strives to find optimal solutions for both parties by taking into account the specificities and needs of each. This pragmatic solution requires constant commitment and close collaboration with customers.

An essential step in this collaboration is to personally meet your future contact. This meeting lays the bases of the partnership and should make it possible to agree on the terms of future deliveries. Flexibility must be shown to adapt to the client's particularities, while ensuring that a standard applicable to all is respected. This balance is essential in order to collect reliable data while ensuring low burden of the clients and to guarantee an effective and efficient processing method on the FSO side.

Each company or administration being different, the implementation of this collaboration can be more or less difficult and therefore requires patience and a considerable amount of in-depth work. This perseverance makes it possible to collect employment data quarterly and separately for each workplace in order to guarantee an optimal geographical distribution of employment in Switzerland.

These data are extracted directly from clients' personnel systems, either as aggregates of the facility level or individually in the form of a staff list. It should be noted that currently the profiling has the social insurance numbers (new AHV number) of each employee for 55 per cent of its clients. In order to facilitate consistency work with AHV funds, FSO is working to increase this proportion so that, in the long term, all clients will benefit from individual data with social insurance numbers. As for data transmission, one third of these data are now transmitted via email and two thirds via a secure data transmission channel (SEDEX), which will eventually become the norm for all customers.

Once the data have been received, the first step is to format the different files and make their content plausible. In a second step, a check is carried out with the company's structure in the SBR to ensure that the data collected cover the client's universe. This essential step also allows this structure to be updated in the SBR. Finally, the data for

the quarter under review should be compared with those for the previous quarter in order to detect and explain fluctuations in employment. If this is the case, the customer is contacted to validate these figures. Once these steps have been completed, the data are loaded into the SBR and transmitted to the job statistic and the statistic on the business structure in accordance with the quality control processes established with these partners.

The process described here is the same for all files received, but the processing time varies greatly from one client to another. This is due to the specificities of the various partners but above all to the complexity of their structure. Faced with the requirement to distribute jobs in each workplace and the fact that the majority of clients locate their employees with their own keys, it is necessary to create the links that link these keys with the corresponding SBR number. This approach can be extremely intensive in terms of working time. This applies to many private sector companies but especially to the public sector as a whole.

The profiling method, thanks to close partnership with large companies and administrations established in Switzerland, and consistent long-term work, makes it possible to update the structures of these large units and cover about 30 per cent of total employment in Switzerland.

Synoptic overview of the main features

Coverage of profiling:

> ‣ Enterprise groups with more than 100 employees and 10 local units in private and public sector

Staff of profiling:

> ‣ 9.2 FTE

Goals of profiling:

> ‣ Regular updating of the business register
> ‣ Collection of data for the production of employment statistics

Periodicity:

> ‣ Quarterly

Variables of the enterprise group collected quarterly:

> ‣ Name of enterprise group level
> ‣ Organizational structure of the enterprise group
> ‣ Name of legal unit level
> ‣ Name of local unit level
> ‣ Activity of local unit level
> ‣ Status of local unit level

Variables of employment collected quarterly (data's delivery per employee):

> ‣ Gender
> ‣ Localization
> ‣ Nationality
> ‣ Social security number

Occupancy rate

> ‣ Cross boarder information
> ‣ Legal unit

    ▸   Local unit level

IT system used for profiling:

    ▸   Swiss Business Enterprise Register

    ▸   Internet information (company website for structural information)

    ▸   SharePoint tool for communication with other statistical users

    ▸   SQL programme for data analysis and preparation files for employment statistic producers

    ▸   Quarterly report (documented on share point tool) of profiling on enterprise group

    ▸   Validation for changes of group structure level, legal unit level and local unit level

    ▸   Validation of employment at the level of the group against administrative data on social security insurance

    ▸   Validation of employment at the local unit level

    ▸   Validation of the evolution of employment between each quarter and each year

    ▸   Validation of the number of legal unit and local unit

    ▸   Validation of the consistency on employment statistics at regional level

    ▸   Validation of activity codes and consistency of these inside the enterprise group

    ▸   Metadata shared with other statistical users:

    ▸   Contact person at enterprise group level (sometimes at the legal unit level) for profiling need

    ▸   Contact person at enterprise group level (sometimes at the legal unit level) for other statistics needs

    ▸   Storage of all exchange (reporting of meeting with the group, telephone or email discussion)

Annex F8: Profiling in the United Kingdom of Great Britain and Northern Ireland

Profiling cases are initially determined by the size of the business. They are then are prioritized by referral from other parts of the organization which have identified an issue, or complexity (number of different activities present). Depending on the size and complexity, the case is handled by a senior or junior profiler.

Most cases can be completed by desk profiling, which always involves telephone contact with the business. Only a small number require a company visit: these are determined by the complexity of the case. It is also important to remember that there is no legal requirement for businesses to participate in the profiling process, but they will benefit from clearer reporting requirements as a result.

An outline of the United Kingdom profiling process is set out below:

Preparatory checks

    ▸   Check the last profile report, check the survey comments database for any relevant information about contact with the business, and check the survey response history on the Inter-Departmental Business Register (IDBR).

➤ Check the company website for background and structural information.

Start profile - using a profiling template

A standard template to maintain consistency and quality of the profile is used in every case. A set of checks are completed and all findings are documented:

Enterprise group checks

➤ Check group employment against administrative data on payroll.

➤ Check for changes of ownership, check on Dun & Bradstreet (D&B) system that the references and ownership structure is correct.

➤ Compare subsidiary data between D&B and IDBR.

Legal unit checks

➤ Check the company numbers, the company name and activity codes (i.e. is it trading, dormant, etc.).

Enterprise and reporting unit checks

➤ Various checks around consistency of the enterprise, reporting units and their associated local units. Checks of classification and employment.

Survey selections

➤ Check and ensure that all contacts are still valid.

Check local unit

➤ Check for duplication, check names and check business description against Standard Industrial Classification (NACE).

Turnover congruence

➤ Check turnover of the units, with other sources such as survey and administrative data.

Employment congruence

➤ Check employment compared to administrative and other survey sources.

Discussion with the company

➤ Either by telephone or a visit, discuss issues of concern identified above with the company to find and explanation.

Actioning and documentation

➤ Complete actions identified above on the IDBR database.

➤ Complete the profile report and attach this to the survey management database, and the profiling case log.

➤ Set the date of profile completion.

Glossary[89]

Active unit

A unit is active when it has any economic activity or when it has no economic activity but is legally or administratively registered and part of another unit that has economic activity at any time during a respective reference period.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Activity, economic activity.

Activity

An activity is a process, i.e. a combination of actions that result in a certain set of products. Activities are defined as the use of inputs (e.g. capital, labour, energy, and materials) to produce outputs. The outputs that result from undertaking activities can be transferred or sold to other units (in market or non-market transactions), placed in inventory or used by the producing units for own final use.

In practice the majority of units carry on activities of a mixed character. One can distinguish between three types of economic activity:

> ▸ Principal activity: The principal activity is the activity which contributes most to the total value added of the unit under consideration.

> ▸ Secondary activity: A secondary activity is any other activity of the unit that produces goods or services.

> ▸ Ancillary activity: Ancillary activities are any activities that exist solely to support the main productive activities of a unit by providing non-durable goods or services for the use of that entity.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/class ifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Active unit, economic activity, classification of activities, ancillary activity, secondary activity, principal activity.

Administrative business register

An administrative business register is a regularly updated structured list of specific business units in a territorial area, which is maintained by administrative authorities for administrative, legal or taxation purposes (e.g. recording and maintaining certain details of businesses or taxation).

Related terms: Administrative register, statistical business register.

Administrative data

Administrative data are data originally collected for non-statistical purpose. Control of the methods by which the administrative data are collected and processed rests with the administrative agency. In most cases the administrative authority will be a government unit.

Source: United Nations Economic Commission for Europe, "Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices", United Nations, New York and Geneva, 2011.

Link: www.unece.org/fileadmin/DAM/stats/publications/Using_Administrative_Sources_Final_for_web.pdf

Related terms: Administrative register, administrative source.

Administrative register

Administrative register is an umbrella term and covers for example records collected for business registers or personal registers for administrative purposes. A register is a written and complete record containing regular entries of items and details on particular set of objects. Typically, a register is a structured list of units, containing a number of attributes for each of those units, and having some sort of regular updating mechanism. Registers maintained by administrative authorities for administrative purposes can be considered to be administrative registers..

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Administrative business register.

Administrative source

Administrative sources are files of data collected by government bodies for the purposes of administering taxes and benefits or monitoring populations. More generally, administrative sources contain information that is not primarily collected for statistical purposes.

Source: United Nations Economic Commission for Europe, "Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices", United Nations, New York and Geneva, 2011.

Link: www..unece.org/fileadmin/DAM/stats/publications/Using_Administrative_Sources_Final_for_web.pdf

Related terms: Administrative register, administrative data.

Administrative unit

An administrative unit is designed for the purposes of conforming with an administrative regulation, for example for registration purposes or for accounting purposes of VAT and other taxes.

Related terms: Statistical unit.

All-resident enterprise group

An all-resident enterprise group is an enterprise group composed only of enterprises that are all resident in the same country.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Enterprise group, domestically controlled enterprise group.

Ancillary activity

Ancillary activities are undertaken to support principal and secondary productive activities of a unit by providing goods or services entirely or primarily for the use of that entity, such as bookkeeping, transportation, storage, purchasing, sales promotion, cleaning, repair and maintenance, security, etc. The output is always intended for intermediate consumption within the same unit and is therefore usually not recorded separately. Although most ancillary activities produce services, some goods-producing activities may, by exception, be regarded as ancillary. The goods thus produced, however, may not become a physical part of the output of the main productive activities. Ancillary activities are usually fairly small-scale compared with the principal activity they support.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Activity, economic activity, principal activity, secondary activity.

Attribute

See: Characteristic.

Birth (of enterprise)

The birth of an enterprise is characterized by the creation of a combination of production factors with the restriction that no other enterprises are involved in the event. Births do not include creation of entries into the population due to mergers, break-ups, split-off or restructuring of a set of enterprises. It does not include entries into a subpopulation resulting only from a change of activity. A birth means that the enterprise starts from scratch and actually starts activity. An enterprise creation can be considered as an enterprise birth if new production factors, in particular new jobs, are created. If a dormant unit is reactivated within two years, this event is not considered a birth.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Break-up (of enterprise), creation (of business), date of creation (of enterprise), birith of enterprise, creation (of business, split-off (of enterprise).

Birth (of enterprise group)

The birth of an enterprise group is the establishing of a link of control, direct or indirect, between two or more independent legal units, where no link of control existed before and no other enterprise group is involved.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Break-up (of enterprise group), creation (of business), date of creation (of enterprise group).

Branch

A branch is an unincorporated enterprise that wholly belongs to a non-resident unit, known as the parent. It is resident and treated as a quasi-corporation. For the purpose of the statistical business register a branch shall be treated as an enterprise.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Foreign affiliate, subsidiary.

Break-up (of enterprise)

The break-up of an enterprise results in one enterprise before and more than one enterprise after the event. In a break-up, the enterprise is divided in such a way that neither (none) of the new enterprises keeps the identity of the original enterprise. There is no continuity or survival, but the closure of the previous enterprise is not considered to be a real death. Similarly, the new enterprises are not considered to be real births. A break-up is similar to a split-off and can be seen as the opposite of a merger.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Birth (of enterprise), creation (of business), birth of enterprise, creation (of business), split-off (of enterprise).

Break-up (of enterprise group)

The break-up of an enterprise group results in one enterprise group before and more than one enterprise group after the event. In a break-up, the enterprise group is divided in such a way that neither (none) of the new enterprise groups keeps the identity of the original enterprise group.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Birth (of enterprise group), creation (of business).

Business

Business describes a type of enterprise, namely a commercial enterprise or legal unit with commercial economic activity.

Related terms: Enterprise.

Business closures

See: Cessation of business.

Business demography

Business demography covers events, such as births and other creations of units, deaths and other cessations of units, and their ratio to the business population. It covers follow-up of units in time dimension, thus gaining information on their survival or discontinuity. It also covers the development in time dimension according to certain characteristics like size, thus gaining information on the growth of units, or a cohort of units, by type of activity. Demographic information can in principle be produced for any statistical unit; however, a clear political interest in Europe is on enterprise demography. In other regions, business demography data are often calculated based on establishments. The demography of enterprises can be assessed by studying enterprise births and enterprise deaths and by examining the change in the number of enterprises by type of activity, i.e. by examining the flows and stocks to get a complete picture of the enterprise dynamism.

Source: European Commission (Eurostat), OECD, *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working papers, Publication Office of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

19.1. **Related terms:** Continuity (of enterprise, survival.

Business start-ups

See: Creation (of business).

Business statistics

See: Economic statistics.

Captive financial institution

Captive financial institutions are SPEs that either simply own the assets of subsidiary corporations, or are units with the characteristics of SPEs, including investment and pension funds and units used for holding and managing wealth for individuals or families, holding assets for securitization, issuing debt securities on behalf of related companies (such a company may be called a conduit), securitization vehicles and carrying out other financial functions. The degree of independence from its parent may be demonstrated by exercising some substantive control over its assets and liabilities

to the extent of carrying the risks and reaping the rewards associated with the assets and liabilities.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Control, ownership, special purpose entity.

Cessation (of business)

The cessation of business can occur either due to a (real) death of a unit, or due to other cessation by a merger, takeover, break-up or discontinuity point according to the continuity rules.

Source: European Commission (Eurostat), OECD, *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working papers, Publication Office of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Death (of enterprise), death (of enterprise group).

Characteristic

A characteristic is one element of a set of information that is stored in a business register to describe a statistical unit. Characteristics are provided for identification of a unit, such as name, address, and identification numbers, for economic description of a unit, such as activity code, turnover or employment of a unit, or for the structure of a unit, such as the relationship to other statistical units.

Related terms: Variable.

Classification of activities

A classification of activities has as its main purpose to classify economic entities and statistical units, such as establishments or enterprises, according to the economic activity in which they mainly engage. The main aim is to provide a set of activity categories that can be utilized when dissecting statistics according to such activities. Different classifications are needed to cater for the different functions which statistics are required to perform, so at international and national levels classifications have been developed for a wide range of purposes, whereby each has its own specific area of application depending on the subject of classification. Economic classifications can be broadly divided into two categories:

Classifications of economic activities, which aim to cover all economic activities – from agriculture to services – are used to classify economic entities (enterprises, establishments, local units or other statistical units). Such classifications therefore form the basis for compiling statistics on output, the production factors entering into the production process (input: labour, raw materials and supplies, energy, etc.), capital formation or financial transactions. The international classification for activity is the ISIC, maintained by the United Nations, used at world level.

Classifications of products. The outputs of the economic entities are termed products and are generally divided into goods and services and grouped according to a product classification. The international classification for products is the Central Product Classification (CPC), maintained by the United Nations and used worldwide.

Source: United Nations, International Standard Industrial Classification of all Economic Activities (ISIC), Rev. 4, 2008; and CPC Version 2.1, 2015.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_
Rev_4_publication_English.pdf; https://unstats.un.org/unsd/classifications/Econ/
Download/In%20Text/CPCv2.1_complete(PDF)_English.pdf

Related terms: Activity.

Common frame

See: Frozen frame.

Company

See: Corporation.

Concentration of enterprises

Concentration of enterprises refers to demographic events involving more than one
enterprise before and one enterprise after such events as merging or taking-over. The
term may also be used to denote that the population of enterprises gets fewer owners
or is spread over a reduced number of enterprise groups.

Source: European Commission, Eurostat, "Business Registers Recommendations
Manual", Methodologies and Working papers, Publication Office of the European
Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.
pdf

Related terms: Merger, takeover.

Continuation

See: Survival.

Continuity (of enterprise)

The continuity of an enterprise depends, in principle, on the continuity of its produc-
tion factors: employment, machines and equipment, land, buildings, management,
and intangible assets. In theory, the continuity rules would be derived from the defi-
nition of the enterprise (or other units) and its statistical uses. The continuity of the
production factors can be measured and weighted to decide upon the continuity of the
enterprise. In practice, the continuity rules consider three main criteria: continuity of
control, economic activity and location.

Source: Eurostat, "Business registers. Recommendations manual", Methodologies and
Working Papers, Publications Office of the European Union, Luxembourg, 2010

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.
pdf

Related terms: Business demography, survival.

Control

Control describes the case where a single institutional unit owning more than a half
of the shares, or equity, of a corporation is able to control its policy and operations by
outvoting all other shareholders, if necessary. Similarly, a small, organized group of

shareholders whose combined ownership of shares exceeds 50 per cent of the total is able to control the corporation by acting in concert.

Corporation B is said to be a subsidiary of corporation A when: either corporation A controls more than half of the shareholders' voting power in corporation B; or corporation A is a shareholder in corporation B with the right to appoint or remove a majority of the directors of corporation B. Corporation A may be described as the parent corporation in this situation. As the relationship of a parent corporation to a subsidiary is defined in terms of control rather than ownership, the relationship must be transitive: that is, if C is a subsidiary of B and B is a subsidiary of A, then C must also be a subsidiary of A. If A has a majority shareholding in B while B has a majority shareholding in C, A cannot also have a majority shareholding in C. Nevertheless, A must be able to control C if it controls B.

Source: 2008 SNA

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Ownership, subsidiary.

Corporation

Corporations are institutional units that are capable of generating a profit or other financial gain for their owners, recognized at law as separate legal entities from their owners who enjoy limited liability, and set up for purposes of engaging in market production. Corporations may be described, in the legal sense, by a variety of names: corporations, incorporated enterprises, public limited companies, public corporations, private companies, joint-stock companies, limited liability companies, limited liability partnerships, and so on. In the SNA, the term corporation covers legally constituted corporations and also cooperatives, limited liability partnerships, notional resident units and quasi-corporations.

Source: 2008 SNA

Link:http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Enterprise, enterprise group, establishment, legal unit.

Creation (of business)

Creation refers to the emergence of a new business unit. This can be either due to a (real) birth of the unit, or due to other creation by a merger, break-up, split-off or discontinuity point according to the continuity rules.

Source: European Commission (Eurostat), OECD, *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working papers, Publication Office of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Birth (of enterprise), birth (of enterprise group), break-up (of enterprise), birth of enterprise, creation (of business), takeover.

Date of birth (of enterprise)

See: Date of creation (of enterprise).

Date of birth (of enterprise group)

See: Date of creation (of enterprise).

Date of birth (of legal unit)

See: Date of creation (of legal unit).

Date of birth (of local unit)

See: Date of creation (of local unit).

Date of cessation (of enterprise)

Date of cessation of an enterprise is the date of final cessation of its activities. It refers to the death or other deletion date of the enterprise (when it becomes historical) and is interpreted in a way similar to the corresponding variable for local units. A death amounts to the dissolution of a combination of production factors with the restriction that no other enterprises are involved in the event. Deaths do not include exits from the population due to mergers, takeovers, break-ups or restructuring of a set of enterprises. It does not include exits from a subpopulation resulting only from a change of activity. An enterprise is included in the count of deaths only if it is not reactivated within two years. Equally, a reactivation within two years is not counted as a birth.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN. pdf

Related terms: Cessation (of business), death (of enterprise).

Date of cessation (of enterprise group)

Date of cessation of an enterprise group refers to the date of cessation of the all-resident or truncated enterprise group. Cessation of a group means either death of the group (dissolution of the links of control between the units belonging to the group), or (more commonly) other cessation date by merger with or takeover by another group, or break-up, split-off, or restructure into two or more groups. The death of an enterprise group is the cessation of all control links, direct or indirect, between the legal units of which the enterprise group consists. The legal units become independent again or cease to exist. No other enterprise group is involved.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN. pdf

Related terms: Cessation (of business), death (of enterprise group).

Date of cessation (of legal unit)

Date of cessation of a legal unit refers to the date on which the legal unit ceased to be part of an enterprise. This date is not easy to collect but the registration of the event is far more important that the precise day and month of its having taken place. Basically, the legal unit ceases to be part of an enterprise when: The legal unit ceases to exist or the legal unit ceases to be economically active and it is not part of the control chain within the enterprise group.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Cessation (of business), dormant unit.

Date of cessation (of local unit)

Date of cessation of a local unit refers to the date of its final cessation of activities. It refers to the death or other deletion date of the local unit. As for legal units, this date may not be available with any precision, only the fact that the local unit has ceased to exist during the reference year may be known. Since the local unit is a part of an enterprise, situated in a geographically identified place, and the enterprise is a combination of production factors, the death of a local unit amounts to the dissolution of a (partial) combination of production factors at a geographically identified place.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Cessation (of business).

Date of creation (of enterprise)

Date of creation of an enterprise refers to the date on which a unit commenced activities. This can be the date of birth, i.e. in principle the date on which the first financial commitments are made, although in practice it may refer to the registration date in the administrative source, if the unit starts its economic activities immediately after that. However, the legal unit may change and be reregistered for instance after a change of legal form, while the enterprise remains the same, because the continuity rules for enterprises should be applied.

Source: European Commission (Eurostat), OECD, *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working papers, Publication Office of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Birth (of enterprise), creation (of business).

Date of creation (of enterprise group)

Date of creation of an enterprise group refers to the date of commencement of the all-resident or truncated enterprise group. The date refers either to a date when a new all-resident group is born, or other creation date of a new group (by merger, break-up, split-off, or restructure). The birth of a new group may be difficult to define in practice, if the smallest groups of no statistical importance to the member State are not monitored. The date from which the group is being monitored shall then be used as a proxy. However, the approximate dates are important in order to know from which year a certain multinational group is monitored in different countries.

Source: European Commission (Eurostat), OECD, *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working papers, Publication Office of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Birth (of enterprise group), creation (of business).

Date of creation (of legal unit)

Date of creation of a legal unit refers to the date of incorporation for legal persons or date of official recognition as an economic operator for natural persons. The "date of official recognition" should be the date on which an identification number is given, or the date on which the legal existence was approved, be it a company/trade register number, a VAT number or other.

Source: European Commission (Eurostat), Organisation for Economic Cooperation and Development (OECD), *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working papers, Publication Office of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Creation (of business).

Date of creation (of local unit)

Date of creation of a local unit refers to the date of commencement of the unit's activities. This date should refer to the birth or other creation date of the local unit according to the continuity rules.

Source: European Commission (Eurostat), OECD, *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working papers, Publication Office of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Creation (of business).

Date of death (of enterprise)

See: Date of cessation (of enterprise).

Date of death (of enterprise group)

See: Date of cessation (of enterprise group).

Date of death (of legal unit)

See: Date of cessation (of legal unit).

Date of death (of local unit)

See: Date of cessation (of local unit).

Death (of enterprise)

The death of an enterprise refers to the dissolution of a combination of production factors with the restriction that no other enterprises are involved in the event. Deaths do

not include exits from the population due to mergers, takeovers, break-ups or restructuring of a set of enterprises. It does not include exits from a subpopulation resulting only from a change of activity.

Source: Commission Regulation No. 2700/98 concerning definitions of characteristics for structural business statistics; see *Eurostat – OECD Manual on Business Demography Statistics*, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Cessation (of business), date of cessation (of enterprise).

Death (of enterprise group)

The death of an enterprise group denotes the cessation of all control links, direct or indirect, between the legal units of which the enterprise group consists. The legal units become independent again or cease to exist. No other enterprise group is involved.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Cessation (of business), date of cessation (of enterprise group).

Deconcentration

Deconcentration refers to demographic events (break-ups and split-offs) involving one enterprise before and more than one enterprise after the events. The events of deconcentration mirror those of concentration: the counterpart of the merger is the break-up and the counterpart of the takeover is the split-off.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Break-up, split-off, concentration.

Delineation

The delineation of statistical units is mainly done by considering groupings or divisions of legal units and applying European Union harmonized rules and/or using classifications. Bad delineation may occur when two units are considered as one and vice versa (for instance a cafeteria attached to a supermarket). Another frequent case of difficulties in delineation is encountered with very small enterprises, which represent a total employment less than the FTE of one year.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Standar statistical units, profiling.

Domestically controlled enterprise group

A domestically controlled enterprise group is a multinational group where the global decision centre is in the country compiling the business register.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Enterprise group, foreign controlled enterprise group.

Dormant unit

A dormant unit is one that is legally alive and has legal personality, but does not carry on any activity and has neither employment nor turnover.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Cessation (of business), reactivation.

Economic activity

Economic activity refers to any activity consisting in the offer of goods and services on a given market is an economic activity. In addition, non-market services contributing to GDP, and also the direct and indirect holdings of active legal units are economic activities for the purpose of statistical business registers.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Active unit, activity.

Economic census

Economic census describes a survey conducted on the full set of economic units belonging to a given population.

Source: ECE, "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000.

Link: https://ec.europa.eu/eurostat/ramon/coded_files/UNECE_TERMINOLOGY_STAT_METADATA_2000_EN.pdf

Related terms: Economic survey, economic unit.

Economic operator

See: Economic unit.

Economic organization

See: Economic unit.

Economic producer

See: Economic unit.

Economic production

Economic production refers to an activity carried out under the control and responsibility of an institutional unit that uses inputs of labour, capital, and goods and services to produce outputs of goods or services.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Activity, economic unit.

Economic statistics

Economic statistics describe the activities of economic transactors and the transactions that take place between them. In the real world, economic entities engaged in the production of goods and services vary in their legal, accounting, organizational and operating structures.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Business statistics.

Economic survey

Economic survey refers to an investigation about the characteristics of a given population of economic units by means of collecting data from a sample of that population and estimating their characteristics through the systematic use of statistical methodology.

Source: ECE, "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000.

Link: https://ec.europa.eu/eurostat/ramon/coded_files/UNECE_TERMINOLOGY_STAT_METADATA_2000_EN.pdf

Related terms: Economic census. economic unit.

Economic territory

Economic territory covers a number of concepts. The most commonly used concept of economic territory is the area under the effective economic control of a single government. However, economic territory may be larger or smaller than this, as in a currency or economic union or a part of a country or the world.

The economic territory includes the land area, airspace, territorial waters, including jurisdiction over fishing rights and rights to fuels or minerals. In a maritime territory,

the economic territory includes islands that belong to the territory. The economic territory also includes territorial enclaves in the rest of the world.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Resident unit.

Economic unit

An economic unit is a legal unit, or part of a legal unit, with economic production as defined in the current version of the SNA.

Related terms: Legal unit, economic production, standard statistical unit.

Employees

Employees are persons who, by agreement, work for a resident enterprise and receive a compensation for their labour. The relationship of employer to employee exists when there is an agreement, which may be formal or informal, between the employer and a person, normally entered into voluntarily by both parties, whereby the person works for the employer in return for remuneration in cash or in kind. The measurement is realized by the actual number of persons employed, and number of employees, both as head counts and, in the latter case, also in FTEs defined as total hours actually worked divided by average number of hours actually worked in full-time jobs. The main uses of these characteristics are in stratification for sampling, analysis and dissemination purposes.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Employment, number of employees, number of persons employed.

Employment

Employment includes all persons, both employees and self-employed persons, engaged in some productive activity that is undertaken by a resident enterprise.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Employees, self-employed.

Enterprise

An enterprise is an institutional unit in its capacity as a producer of goods and services. It is an economic transactor with autonomy in respect of financial and investment decision-making, as well as authority and responsibility for allocating resources for the production of goods and services. It may be engaged in one or more productive activities. An enterprise may be a corporation (or quasi-corporation), an NPI or an unincorporated enterprise. The enterprise is the level of statistical unit at which information relating to its transactions, including financial and balance-sheet accounts, are maintained, and from which international transactions, an international investment position (when applicable), consolidated financial position and net worth can be derived.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Corporation, multinational enterprise group, non-profit institution, quasi-corporation, standard statistical unit, unincorporated enterprise.

Enterprise group

An enterprise group is an association of enterprises bound together by legal and/or financial links. A group of enterprises can have more than one decision-making centre, especially for policy on production, sales and profit. It may centralize certain aspects of financial management and taxation. It constitutes an economic unit which is empowered to make choices, in particular concerning the units which it comprises. An enterprise group is a set of enterprises controlled by the group head.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: All-resident enterprise group, multinational enterprise group, truncated enterprise group.

Entrepreneur

An entrepreneur is a person who owns an enterprise and seeks to generate value through the creation or expansion of economic activity, by identifying and exploiting new products, processes or markets.

Source: Defining Entrepreneurial Activity: Definitions Supporting Frameworks for Data Collection (OECD, 2008)

Link: www.oecd.org/sdd/business-stats/39651330.pdf

Establishment

An establishment is an enterprise or part of an enterprise that is situated in a single location and in which only a single (non-ancillary) productive activity is carried out or in which the principal productive activity accounts for most of the value added.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Standard statistical unit.

EuroGroups Register (EGR)

The EGR builds a framework of registers, consisting of a central register kept at Eurostat and registers in each European Union member state and in EFTA countries. The central register contains information about multinational enterprise groups, which

have statistically relevant financial and non-financial transnational operations in at least one of the European countries. Registers in the European Union member States and in EFTA countries contain information regarding multinational enterprise groups active in the respective countries and are fully consistent with the central register.

The aim of the EGR network is to hold a complete, accurate, consistent and up-to-date set of linked and coordinated statistical registers, which offer compilers a common frame of multinational enterprise groups, global as well as truncated national groups, operating in the economy of the European Union and EFTA countries, together with their constituent legal units and enterprises and the ownership and control relationships between legal units.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: European System of Interoperable Statistical Business Registers, multinational enterprise group.

European System of Accounts (ESA)

ESA 2010 is an internationally compatible accounting framework for a systematic and detailed description of a total economy (that is a region, country or group of countries), its components and its relations with other total economies. ESA 2010 is fully consistent with the worldwide guidelines on national accounting, namely SNA.

Source: European Commission (Eurostat), "European System of National Accounts 2010", Publication Office of the European Union, Luxembourg, 2013.

Link: http://ec.europa.eu/eurostat/documents/3859598/5925693/KS-02-13-269-EN.PDF/

Related terms: System of National Accounts.

European System of Interoperable Statistical Business Registers

The ESBR project (2013–2020) is rationalizing, strengthening and standardizing national SBRs and EGR in the European Statistical System (ESS) with the ultimate aim of making them an efficient interoperable system. It is a continuation of previous EGR and profiling projects carried out in the ESS. A core goal is the improvement of the EGR statistical frames so that they can provide better quality information on multinational enterprise groups for globalization statistics. It includes an updated and agreed methodology for European Union profiling that enables all ESS countries to achieve the same view of multinational enterprises (*seeing the whole elephant);* it is based on a collaborative approach with different countries contributing to profiling the same multinational enterprises according to defined roles and responsibilities and supported by a secure platform for sharing confidential data. The ESBR project also includes practical pilot exercises aimed at testing the proposed solutions with ESS countries.

Source: European Commission (Eurostat), "The European System of Interoperable Business Registers (ESBRs)", Compact guides, Publication Office of the European Union, Luxembourg, 2013.

Link: http://ec.europa.eu/eurostat/documents/4031688/5931580/KS-03-13-411-EN.PDF, https://ec.europa.eu/eurostat/web/ess/esbr

Related terms: EuroGroups Register.

Financial corporation

See: Institutional sector.

Firm

Firm usually refers to a market business unit. It is often used as synonym for company or unincorporated business.

Related terms: Business, Company.

Foreign affiliate

Foreign affiliate refers to an enterprise resident in the compiling country over which an institutional unit not resident in the compiling country has control, or an enterprise not resident in the compiling country over which an institutional unit resident in the compiling country has control.

Source: Eurostat, "Foreign AffiliaTes Statistics (FATS) Recommendations Manual", Methodologies and Working papers, 2012 edition

Link: ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/KS-RA-12-016

Related terms: Foreign affiliate statistics (FATS), foreign controlled enterprise group.

Foreign affiliate statistics (FATS)

FATS are statistics describing the overall activity of foreign affiliates. As there are two points of view with regard to the location of the affiliate, FATS are divided into two sets of statistics, inward and outward FATS: inward statistics on foreign affiliates describe the activity of foreign affiliates resident in the compiling economy, while outward statistics on foreign affiliates describe the activity of foreign affiliates abroad controlled by the compiling economy.

Source: Eurostat, "Foreign AffiliaTes Statistics (FATS) Recommendations Manual", Methodologies and Working papers, 2012.

Link: http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-12-016

Related terms: Foreign affiliate, foreign controlled enterprise group.

Foreign-controlled enterprise group

A foreign controlled enterprise group is a multinational group where the global decision centre is outside the country compiling the business register.

Source: Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Enterprise group.

Foreign direct investment (FDI) statistics

FDI statistics embody four distinct statistical accounts: investment positions, financial transactions, associated income flows between enterprises that are related through

a direct investment relationship, and other changes in the value of assets, especially revaluation terms. Direct investment is a category of cross-border investment associated with a resident in one economy (the direct investor) having control or a significant degree of influence on the management of an enterprise (the direct investment enterprise) that is resident in another economy.

Source: OECD, "OECD Benchmark Definition of Foreign Direct Investment", fourth edition, 2008.

Link: http://unstats.un.org/unsd/EconStatKB/Attachment549.aspx

Related terms: Inward foreign affiliate statistics (I-FATS), outward foreign affiliate statistics (O-FATS).

Frame

The frame for a given survey is the subset of the frozen frame, comprising the set of units that match the specification of the survey target population. Thus, for example, a survey of employment will include units in all (or at least most) industries that are employers, i.e. will exclude units that are non-employers. A survey of manufacturing will include all units that have an ISIC code in the manufacturing group, whether they have employees or not. A survey of capital expenditure may include all units above a certain size. Thus, the survey frames are typically different from one another but may be extracted from the same common set of units, i.e. a frozen frame. A frame may be referred to as a survey frame, and also as a sampling frame or a survey sampling frame.

Source: United Nations Economic Commission for Europe, Guidelines on Statistical Business Registers, Geneva, 2015.

Link: https://unece.org/fileadmin/DAM/stats/publications/2015/ECE_CES_39_WEB.pdf

Related terms: Frozen frame.

Frozen frame

The frozen frame is a subset of the snapshot that comprises all statistical units that are active, or potentially active, or active within the previous reference year. It also includes administrative units that are linked to these statistical units. The aim is to include all units and all characteristics that are used by subsequent processes. In other words, it is a trimmed down version of the snapshot that is easier to manipulate because the possible large number of inactive units are not there. It may be further restricted by containing only units for which there are values for the characteristics that are to be used for frame extraction and sample selection for at least one survey.

Source: United Nations Economic Commission for Europe, Guidelines on Statistical Business Registers, Geneva, 2015

Link: https://unece.org/fileadmin/DAM/stats/publications/2015/ECE_CES_39_WEB.pdf

Related terms: Frame, live register.

Frozen register

See: Register snapshot.

General government

See: Institutional sector.

Global decision centre

The global decision centre is the unit where the strategic decisions referring to an enterprise group are taken.

Source: Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Global group head, multinational enterprise group.

Global value chain (GVC)

GVC describes the full range of activities that firms and workers perform, on a global scale, to bring a product from its conception to end use and beyond. This includes activities such as research and development, design, production, marketing, distribution and support to the final consumer.

Source: Gereffi, Gary and Fernandez-Stark, Karina, Global Value Chain Analysis: A Primer, Second Edition, The Duke Center on Globalization, Governance & Competitiveness, July 2016.

Link: https://unstats.un.org/unsd/business-stat/GVC/

Government unit

Government units are unique kinds of legal entities established by political processes that have legislative, judicial or executive authority over other institutional units within a given area.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Institutional unit.

Group head

The group head is a parent legal unit of an enterprise group, which is not controlled either directly or indirectly by any other legal unit. In the case of multinational enterprise groups global and domestic group heads can be identified. The global group head is the group head of the multinational enterprise group, the domestic group head is on the top of the truncated national part of the multinational enterprise group.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Global decision centre, multinational enterprise group.

GSBPM

GSBPM models the phases of the statistical business process and provides generic terms to describe them. GSBPM is used to harmonize statistical computing infrastructures, facilitate the sharing of software components and provide a framework for process quality assessment and improvement. GSBPM is intended to apply to all activities undertaken by producers of official statistics, at both the national and international levels, that result in data outputs. It is designed to be independent of the data source,

so that it can be used for the description and quality assessment of processes based on surveys, censuses, administrative records, and other non-statistical or mixed sources.

Source: ECE, UNECE Statistics Wiki. As of 1 July 2019, the current version is GSBPM v5.1.

Link: https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model

Related terms: GSIM.

GSIM

GSIM is a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information that are used in the production of official statistics (information objects). Examples of information objects in scope for GSIM include data (held in data sets) and metadata (e.g. classifications, variables, questions and populations).

Source: ECE, Statistical Metadata Wiki (METIS-Wiki), last consulted on 1 May 2015.

Link: www.1.unece.org/stat/platform/display/gsim/Brochures

Related terms: GSBPM.

Head office

Head offices are units exercising some aspects of managerial control over their subsidiaries. Their activities include the overseeing and managing of other units of the company or enterprise; undertaking the strategic or organizational planning and decision-making role of the company or enterprise; exercising operational control and managing the day-to-day operations of their related units.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Holding company.

Historical data

Historical data come into existence when a unit has ceased and is not dormant (temporary inactive). For the purpose of reconstructing demographic events the record needs to be marked historical and not to be deleted physically.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Cessation (of business), historical register.

Historical register

A historical register is a version of the register with the capacity to view the content of a live register at some time point in the past.

Related terms: Historical data, live register.

Holding company

A holding company is a unit that holds the assets of subsidiary corporations but does not undertake any management activities. The principal activity of a holding is owning the group. The holding company does not provide any other service to the enterprises in which the equity is held, i.e. it does not administer or manage other units.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Head office.

Horizontal integration

Horizontal integration occurs when an activity results in end-products with different characteristics. This could theoretically be interpreted as activities carried out simultaneously using the same factors of production, in which case it would often be impossible to separate such activities statistically into different processes, assign them to different units or generally provide separate data for them, nor would rules relying on allocation of value added or similar measures be applicable. Alternative indicators, such as gross output, might sometimes be applicable, but there is no general rule for identifying the single activity that best represents the mix included in this horizontal integration. Since patterns of horizontal integration have been considered in the preparation of the classification, in many cases commonly integrated activities are included in the same class of activity even though their outputs have quite different characters.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Activity, vertical integration.

Household

See: Institutional sector.

Identifier

An identifier has the purpose of identifying a unit and linking it with other units in the register and with administrative and statistical sources. The identity number of a legal unit can be either specific to the SBR or an external one, common or shared with other institutions in the Member State, a so-called unique identifier.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Characteristic.

Indirect control

Indirect control means that a parent unit has control over a subunit (sub-subsidiary) through one or many other subsidiaries. Indirect control does not require the parent

unit to own a majority of an integrated shareholding in the capital share of the sub-subsidiaries.

Source: European Commission, Eurostat, "Foreign AffiliaTes Statistics (FATS) Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2012.

Link: http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-12-016

Related terms: Control, ownership, parent corporation.

Industry

An industry consists of a group of establishments engaged in the same, or similar, kinds of activity. At the most detailed level of classification, an industry consists of all the establishments falling within a single ISIC class. At higher levels of aggregation corresponding to the groups, divisions and, ultimately, sections of the ISIC, industries consist of a number of establishments engaged in similar types of activities.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Establishment, kind-of-activity.

Informal sector

The informal sector is broadly characterized as consisting of units engaged in the production of goods or services with the primary objective of generating employment and incomes to the persons concerned. These units typically operate at a low level of organization, with little or no division between labour and capital as factors of production and on a small scale. Labour relations – where they exist – are based mostly on casual employment, kinship or personal and social relations rather than contractual arrangements with formal guarantees. The informal sector thus defined excludes households producing exclusively for own final use.

Source: Measuring the Non-Observed Economy: A Handbook, OECD, IMF, ILO, Interstate Statistical Committee of the Commonwealth of Independent States, 2002.

Link: www.oecd.org/dataoecd/9/20/1963116.pdf

Institutional sector

Institutional sector refers to a group of institutional units that have similar principal functions, behaviour and objectives. There are five institutional sectors in the 2008 SNA: non-financial corporations, financial corporations. federal Government, NPISH.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Institutional unit.

Institutional unit

An institutional unit is an economic unit that is capable, in its own right, of owning assets, incurring liabilities and engaging in economic activities and in transactions

with other entities. Thus an institutional unit is entitled to own goods or assets in its own right; to exchange ownership of goods or assets in transactions with other institutional units, is able to take economic decisions and engage in economic activities for which it is itself held to be directly responsible and accountable at law, is able to incur liabilities on its own behalf, to take on other obligations or future commitments and to enter into contracts, has a complete set of accounts or it would be possible to compile a complete set of accounts if they were required.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Institutional sector, standard statistical unit.

Inward foreign affiliates statistics (I-FATS)

Inward statistics on foreign affiliates describe the activity of foreign affiliates resident in the compiling economy.

Source: European Commission, Eurostat, "Foreign AffiliaTes Statistics (FATS) Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2012.

Link: http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-12-016

Related terms: Foreign affliate statistics (FATS), foreign direct investment (FDI) statistics, outward foreign affiliate statistics (O-FATS).

Kind-of-activity unit (KAU)

A KAU is an enterprise or part of an enterprise that engages in only one kind of productive activity or in which the principal productive activity accounts for most of the value added. Compared with the establishment, in the case of such a unit, there is no restriction on the geographical area in which the activity is carried out but it is characterized by homogeneity of activity. Compared with the establishment, on the geographic area in which the activity is carried out but it is characterized by homogeneity of activity.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Establishment

Legal form

The legal form (also known as legal status) is defined according to national legislation. It is useful for eliminating ambiguity in identification searches and as the possible criterion for selection or stratification for surveys. It is also used for defining the institutional sector. Statistics according to legal form are produced, for example, in business demography. The character of legal or natural person is decisive in fiscal terms, because the tax regime applicable to the unit depends on this. It means that any statistical register fed with fiscal records will have that information. Experience has shown that legal form will often be useful when making adjustments to information collection processes and questionnaires on the legal unit operating an enterprise. A code representing the legal form should therefore be recorded in accordance with the

classification of legal forms or categories. The following legal forms can be found in most countries:

> ▸ Sole proprietorship: Enterprise owned exclusively by one natural person.

> ▸ Partnership: Association of persons who conduct a business under a collective name. It can take the form of a limited partnership.

> ▸ Limited liability companies: Enterprises comprising joint-stock companies, limited partnerships with share capital and private limited company.

> ▸ Cooperative societies: These are bodies set down by law in each country. They observe a number of general principles, for example they may only be entitled to provide their services to members, profits are often distributed in proportion to members' dealings with the society, etc.

> ▸ Non-profit institutions.

> ▸ Enterprises with other forms of legal constitution: This group includes non-taxative nationalized industries in form of publicly-owned enterprises and state or local authority monopolies.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Legal person, legal unit.

Legal person

Legal person refers to all forms of legal construction organized by the constitution and laws of countries and endowed with rights and obligations characteristic of legal personality.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Legal form, legal unit, natural person.

Legal unit

Legal units include:

> ▸ Legal persons whose existence is recognized by law independently of the individuals or institutions which may own them or are members of them.

> ▸ Natural persons who are engaged in an economic activity in their own right.

The legal unit is usually recorded in one or more administrative sources. The sources used for SBRs do not necessarily provide identical views of legal units. These units may vary both between different sources within a country and between countries. Thus, the legal unit is not suitable as a statistical unit, particularly for international comparisons. The characteristics of a legal unit are: it owns goods or assets, it incurs liabilities and it enters into contracts. The legal unit always forms, either by itself or sometimes in combination with other legal units, the basis for the statistical unit known as the "enterprise".

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Legal form, legal person, natural person.

Live register

The live register is the part of an SBR that is being continually updated with new information about the units and their characteristics.

Related terms: Administrative business register, EuroGroups Register, statistical business register.

Local kind-of-activity unit (LKAU)

See: Establishment.

Local unit

A local unit is an enterprise or part of an enterprise (e.g. a workshop, factory, warehouse, office, mine or depot) that is engaged in productive activity at or from one location. The definition has only one dimension in that it does not refer to the kind of activity that is carried out.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Enterprise, standard statistical unit.

Majority ownership

Majority ownership describes the case where a single investor controls an enterprise by holding a majority (more than 50 per cent) of the voting power or of the shares, directly or indirectly. While majority ownership is the major criterion for determining control, it is not indispensable for exercising control. A government can exert control through a legislative decree or regulation, empowering it (the government) to determine the enterprise's policy or to appoint (a majority of) directors.

Source: European Commission, Eurostat, "Foreign AffiliaTes Statistics (FATS) Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2012.

Link: http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-12-016

Related terms: Control, ownership.

Market producer

A market producer is a corporation that is created for the purpose of producing goods or services for sale on the market at prices that are economically significant.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Non-market output, production unit.

Master frame

See: Frozen frame.

Merger

A merger occurs when enterprises integrate and the number of existing enterprises is reduced. If two enterprises integrate entirely, the enterprises involved may lose their identity because they are dissolved beyond recognition in the new organization. If both enterprises lose their identity, the event is called a merger. There is no continuity or survival, but the closures of the previous enterprises are not considered to be real deaths. Similarly the new enterprise is not considered to be a real birth. This event can be seen as the opposite of a break-up.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: enterprise, take-over (of enterprise).

Multinational enterprise group

A multinational enterprise group is an enterprise group which has at least two enterprises or legal units located in different countries.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: All-resident enterprise group, enterprise group, truncated enterprise group.

Natural person

The term natural person is used by the law and by many administrative authorities to denote a human being endowed with all the rights constituting legal personality.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Legal person.

Nature-of-business survey

See: SBR improvement survey.

New enterprise

See: Birth (of enterprise), creation (of business).

Non-financial corporation

See: Institutional sector.

Non-market output

Non-market output consists of goods and individual or collective services produced by NPISHs or government that are supplied free, or at prices that are not economically significant, to other institutional units or the community as a whole. A price is said to be not economically significant when it has little or no influence on how much the producer is prepared to supply and is expected to have only a marginal influence on the quantities demanded.

Source: 2008 SNA

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Market producer.

Non-observed economy

Non-observed economy describes activities that, for one reason or another, are not captured in regular statistical enquiries. The reason may be that the activity is informal and thus escapes the attention of surveys geared to formal activities, the producer may be anxious to conceal a legal activity, or it may be that the activity is illegal.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Informal sector, Market producer.

Non-profit institution

> ‣   Non-profit institutions (NPIs) are legal or social entities created for the purpose of producing goods and services but whose status does not permit them to be a source of income, profit or other financial gain for the units that establish, control or finance them.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Non-market output, NPISH.

Non-profit institution serving households (NPISH)

See: Institutional sector.

Number of employees

Number of employees refers to those persons who work for an employer and who have a contract of employment and receive compensation in the form of wages, salaries, fees, gratuities, piecework pay or remuneration in kind. A worker from an employment agency is considered to be an employee of that employment agency and not of the units in which they (temporarily) work.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Employees, employment, number of persons employed.

Number of persons employed

Number of persons employed refers to the total number of persons who work in the unit, including wageearners and self-employed persons (that is, working proprietors, partners working regularly in the unit and unpaid family workers), as well as persons who work outside the unit but who belong to it and are paid by it (such as sales representatives, delivery personnel, repair and maintenance teams). It excludes labour supplied to the unit by other enterprises, persons carrying out repair and maintenance work in the unit on behalf of other enterprises, as well as those on compulsory military service.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Employees, employment, number of employees.

Observation unit

Observation unit is a unit about which data are obtained during the course of a survey. It is usually a statistical unit, or, if data cannot be obtained about a statistical unit, then some other unit about which data can be obtained and from which data for a statistical unit can be compiled.

Related terms: Reporting unit, standard statistical unit.

Offshoring

Offshoring is generally defined as companies' purchases of intermediate goods and services from foreign providers at arm's length or the transfer of particular tasks within the firm to a foreign location, that is, to foreign affiliates.

Source: OECD, "Measuring Globalization: OECD Economic Globalization Indicators 2010", OECD, 2010.

Link: www.oecd-ilibrary.org/docserver/download/9210031e.pdf?expires=1421077
118&id=id&accname=ocid194935&checksum=9AC2705CA7526D789461FA65C00
2DF58

Related terms: Outsourcing.

Operational unit

Operational unit is a unit defined by a legal unit for the purposes of organizing itself, for example a division, branch, workshop, warehouse, or outlet. In small enterprises, the operational and legal structures often coincide and may even be embodied in a single unit. For large enterprises, the operational structure may be different from the legal structure, coinciding with it only at the highest level of the business. In such cases, the organizational and production units of the enterprise's operational structure may differ from the units of their legal structure. The statistical units of large and complex institutional units may be delineated through a process referred to as profiling. Profiling identifies the enterprise, its legal structure, its operating structure, and the production and organizational units that are used to derive the statistical units.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_
Rev_4_publication_English.pdf

Related terms: Legal unit.

Outsourcing

Outsourcing refers to the purchasing of intermediate goods and services from outside specialist providers at arm's length either nationally or internationally.

Source: OECD, "Measuring Globalization: OECD Economic Globalization Indicators 2010", OECD, 2010.

Link: www.oecd-ilibrary.org/docserver/download/9210031e.pdf?expires=1421077118& id=id&accname=ocid194935&checksum=9AC2705CA7526D789461FA65C002DF58

Related terms: Offshoring.

Outward foreign affiliates statistics (O-FATS)

Outward statistics on foreign affiliates describe the activity of foreign affiliates abroad controlled by the compiling economy.

Source: European Commission, Eurostat, "Foreign AffiliaTes Statistics (FATS) Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2012.

Link: http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-12-016

Related terms: Foreign affliate statistics (FATS), foreign direct investment (FDI) statistics, inward foreign affiliate statistics (I-FATS).

Ownership

Ownership of a unit or a group of units is related to the property of its assets and determines the distribution of financial flows and income. If a unit or group of units is owned by shareholders, its ownership is vested in the shareholders collectively and can be seen as diffused among the legal units that own its shares in proportion to their shareholdings, and independently of voting rights.

Source: Eurostat, "Business registers. Recommendations manual", Methodologies and Working Papers, Publications Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Control, parent corporation, subsidiary.

Parent corporation

See: Control.

Principal activity

The principal activity is the activity which contributes most to the total value added of the unit under consideration. Ideally, the principal activity of the unit should be determined with reference to the value added to the goods and services produced, by applying the top-down method. The top-down method follows a hierarchical principle: the classification of the unit at the lowest level of the classification must be consistent with the classification of the unit at the higher levels. The principal activity so identified does not necessarily account for 50 per cent or more of the entity's total value added. In case that the value added is not known, substitute criteria can be used as proxies to

obtain the best approximation possible compared to the result which would have been obtained on the basis of value added data.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Activity, economic activity, ancillary activity, secondary activity.

Producer

See: Production unit.

Producing unit

See: Production unit.

Product

Products are goods and services (including knowledge-capturing products) that result from a process of production. They are exchanged and used for various purposes: as inputs in the production of other goods and services, as final consumption or for investment. The SNA makes a conceptual distinction between market, own final use and non-market goods and services, allowing in principle any kind of good or service to be any of these three types.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Production unit.

Production unit

A production unit carries out an economic activity under the control and responsibility of an institutional unit using inputs of labour, capital and goods and services to produce outputs of goods and services. Enterprises may be very heterogeneous if they have several secondary activities that are quite different from their principal activities. In order to obtain groups of producers whose activities are more homogeneous, enterprises are partitioned into smaller and more homogeneous units of production like local units, KAUs, and establishments.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Economic unit, institutional unit, market producer, product, standard statistical unit.

Profiling

Profiling is the method used to analyse the legal, operational and accounting structure of an enterprise group at national and world level, in order to establish the statistical units within that group, their links, and the most efficient structures for the collection of statistical data.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Delineation.

Property

See: Characteristic.

Quality

Quality is the degree to which a set of inherent characteristics of an object fulfils requirements as defined in ISO 9000:2015. A simple definition is "fit for use" or "fit for purpose". It is the users' needs that define the quality. Different users may have different needs that must be balanced against each other.

Source: United Nations, Statistics Division, "United Nations National Quality Assurance Frameworks Manual for Official Statistics", Studies in Methods, Series M No. 100, New York, 2019.

Link: https://unstats.un.org/unsd/methodology/dataquality/references/1902216-UNNQAFManual-WEB.pdf

Related terms: Quality indicator.

Quality improvement survey

See: SBR improvement survey.

Quality indicator

Quality indicator is an indicator of the degree of quality of a certain quality aspect.

Source: European Commission, Eurostat, "ESS Handbook for quality Reports", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2009.

Link: http://ec.europa.eu/eurostat/documents/3859598/6651706/KS-GQ-15-003-EN-N.pdf/18dd4bf0-8de6-4f3f-9adb-fab92db1a568

Related terms: Quality.

Quasi-corporations

Quasi-corporations are entities such as certain unincorporated enterprises which may behave in much the same way as corporations. In the SNA such enterprises should be treated as quasi-corporations if they have a complete set of accounts, and included with corporations in the non-financial and financial corporations sectors. SNA recognizes three main kinds of quasi-corporations: first, unincorporated enterprises owned by government units engaged in market production and operated in a similar way to publicly owned corporations; second, unincorporated enterprises, including unincorporated partnerships or trusts, owned by households that are operated as if they were privately owned corporations; and, third, unincorporated enterprises that belong to institutional units resident abroad, referred to as "branches". The SBR should cover quasi-corporations.

Source: 2008 SNA

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Corporations, unincorporated enterprise.

Reactivation

Reactivation describes the event when an enterprise, having been dormant for a period of less than two years, then recommences activity in a way that complies with the definition of continuity. In terms of business demography this event does not constitute a birth or death.

Source: Eurostat and OECD, *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working Papers, Office for Official Publications of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Dormant unit.

Register picture

See: Register snapshot.

Register snapshot

A register snapshot is a copy of all the statistical units in the live register as of a given point in time, including all administrative units or links from statistical units to administrative units. The snapshot is an intermediary step between the live register and the frozen frame. It is used to check for errors that have crept in during processing since the previous snapshot, also as the basis for a historical record. It may contain many inactive enterprises.

Related terms: Live register, frozen frame.

Reporting unit

The reporting unit is the unit that reports to the survey authority. It reports information for each of the observation units. In certain cases it may correspond to an observation unit. An example where it is not the same is where an accounting business reports data on behalf of a client business that is the actual subject of the survey.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Observation unit.

Resident unit

Resident unit is the term used to describe an institutional unit of a country when it has a centre of predominant economic interest in the economic territory of that country: that is, when it engages for an extended period (one year or more being taken as a practical guideline) in economic activities on this territory.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Institutional unit, economic territory.

Restructuring

Restructuring within an enterprise does not affect the continuity of the enterprise, but changes its structure in the process. An example could be the creation or deletion of a local unit. Restructuring may affect key characteristics such as size or principal activity. It could be argued that this is not really a demographic event at the level of the enterprise and does not impact on the demographic variables relating to the enterprise, but it could affect the way the enterprise is included in demographic statistics. Restructuring within an enterprise group is a change (such as the creation or cessation of one or more enterprises) involving more than one enterprise before and more than one enterprise after the event, where all enterprises involved are under common control. It affects the identity of at least one enterprise, though the total number of enterprises before and after the event may be the same. A typical example is the complete reorganization of the production capacity of a large enterprise group. Complex restructuring is a similar event, but this is not constrained to one enterprise group. Restructuring within an enterprise group, or complex restructuring, may entail any number of register creations and deletions.

Source: Eurostat and OECD, "Eurostat-OECD Manual on Business Demography Statistics (Edition 2007)", Methodologies and Working Papers, Office for Official Publications of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Continuity (of enterprise), survival.

Revenue

See: Turnover.

Sampling frame

See: Frame.

Satellite register

A satellite register is the result of a process of extracting parts or the whole of one or more frozen frames from the SBR and then linking information from other data sources to its units. This process is undertaken to extend the functionality of the SBR, which may be a complex network of databases and functionalities, with minimum complication. The responsibility for and control of a satellite register are separated from the SBR and usually take place in a different environment.

Related terms: Statistical business register.

Secondary activity

A secondary activity is each separate activity that produces products eventually for third parties and that is not the principal activity of the unit in question. The outputs of secondary activities are secondary products. Most economic entities produce at least some secondary products.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Activity, economic activity, ancillary activity, principal activity.

Self-employed

Self-employed persons are persons who are the sole or joint owners of the unincorporated enterprises in which they work. Persons who work in unincorporated enterprises are classed as self-employed persons if they are not in paid employment that constitutes their principal source of income; in that latter case, they are classified as employees.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Employment.

Small and medium-sized enterprises (SMEs)

SMEs are non-subsidiary, independent firms which employ fewer than a given number of employees. This number varies across countries. The most frequent upper limit designating a SME is 250 employees, as in the European Union. Some countries, however, set the limit at 200 employees, while the United States considers SMEs to include firms with fewer than 500 employees. Small firms are generally those with fewer than 50 employees, while micro enterprises have at most 10, or in some cases 5, workers. Financial assets are also used to define SMEs. In the European Union, a new definition came into force on 1 January 2005, applicable to all Community acts and funding programmes and also to State aid, whereby SMEs can be granted higher intensity of national and regional aid than large companies. The definition includes financial ceilings: the turnover of medium-sized enterprises (50–249 employees) should not exceed 50 million euros; that of small enterprises (10–49 employees) should not exceed 10 million euros, while that of micro firms (fewer than 10 employees) should not exceed 2 million euros. Alternatively, balance sheets for medium, small and micro enterprises should not exceed 43 million, 10 million and 2 million euros, respectively.

Source: OECD Glossary of statistical terms

Link: http://stats.oecd.org/glossary/

Related terms: Enterprise.

Special purpose entity (SPE)

SPEs do not have a standard definition, but some of the following characteristics may apply. They often have no employees and no non-financial assets. They may have little physical presence beyond a "brass plate" confirming their place of registration. They are always related to another corporation, often as a subsidiary, and are often resident in a territory other than the territory of residence of the related corporations. In the absence of any physical dimension to an enterprise, the residence of an SPE is determined according to the economic territory under whose laws the enterprise is incorporated or registered. Entities of this type are commonly managed by employees of another corporation which may or may not be a related one. The unit pays fees for services rendered to it and in turn charges its parent or other related corporation a fee to cover these costs. This is the only production in which the unit is involved, although it will often incur liabilities on behalf of its owner and will usually receive investment income and holding gains on the assets that it holds. Whether a unit has all or none of these characteristics, and whether or not it is described as an SPE or by some similar designation, it is treated in the SNA in the same way as any other institutional unit by being allocated to sector and industry according to its principal activity unless it falls into one of the three following categories: captive financial institutions, artificial subsidiaries of corporations, or special purpose units of general government.

Source: 2008 SNA

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Institutional unit.

Split-off (of enterprise)

Split-off involves one enterprise before and more than one enterprise after the event. In a split-off, the new enterprise(s) is (are) generally much smaller and the identity of the original enterprise is retained by the larger enterprise. There is no death, but one or more new enterprises are created. This event can be seen as the opposite of a takeover."

Source: Eurostat and OECD, *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working Papers, Office for Official Publications of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

19.2.   **Related terms:** Break-up (of enterprise), birth of enterprise, creation (of business.

Standard statistical unit

Standard statistical units are the units for which information is sought and for which statistics are ultimately compiled. Commonly used types of statistical units for economic statistics are the enterprise, the enterprise group, the KAU, the local unit and the establishment (in Europe called LKAU). In national accounts also the institutional unit is of importance.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Legal unit, institutional unit.

Statistical business register (SBR)

The SBR is a comprehensive, regularly updated and structured list of business units engaged in the production of goods and services, which is maintained by national statistical authorities for statistical purposes to assist in the compilation of statistical data and particular as a (backbone) tool for the preparation and coordination of surveys, as a source of information for statistical analysis of the business population and its demography, for the use of administrative data, and for the identification and construction of statistical units.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Administrative business register.

SBR coverage survey

See: SBR improvement survey.

SBR improvement survey

The SBR improvement survey is a survey conducted by statisticians to improve SBR quality.

Related terms: Statistical business register.

SBR quality improvement survey

See: SBR improvement survey.

SBR survey

See: SBR improvement survey.

Statistical Data and Metadata eXchange (SDMX)

The SDMX initiative sets standards to facilitate the exchange of statistical data and metadata using modern IT. Several versions of the technical specifications have been released since 2004. SDMX has also been published as an ISO International Standard (IS 17369).

SDMX is sponsored by seven international organizations, namely, the Bank of International Settlements, the ECB, Eurostat, the IMF, the OECD, the United Nations Statistics Division and the World Bank.

Link: https://sdmx.org/?page_id=5008

Statistical unit

See: Standard statistical units.

Subcontracting

See: Outsourcing.

Subsidiary

See: Control.

Survey frame

See: Frame.

Survival

Survival refers to the situation where a unit is active and identifiable both before and after a specific (business) demographic event. The unit may be changed in some way, for example, in terms of economic activity, size, ownership or location, but there should be continuity of the unit reference number in the SBR.

In the context of business demography context, survival occurs if an enterprise is active in terms of employment and/or turnover in the year of birth and the following year(s). An enterprise is also considered to have survived if the linked legal units have ceased to be active, but their activity has been taken over by a new legal unit set up specifically to take over the factors of production of that enterprise (this being referred to as survival by takeover).

Source: European Commission (Eurostat), OECD, *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working papers, Publication Office of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Business demography, continuity (of enterprise).

System of National Accounts (SNA)

The System of National Accounts (SNA) is the internationally agreed standard set of recommendations on how to compile measures of economic activity in accordance with strict accounting conventions based on economic principles. The recommendations are expressed in terms of a set of concepts, definitions, classifications and accounting rules that constitute the internationally agreed standard for measuring such items as GDP, the most frequently quoted indicator of economic performance. The accounts present in a condensed way a mass of detailed information, organized according to economic principles and perceptions, about the working of an economy. They provide a comprehensive and detailed record of the complex economic activities taking place within an economy and of the interaction between the different institutional units, and groups of units, which takes place on markets or elsewhere.

The SNA has been prepared under the joint responsibility of the United Nations, IMF, the Commission of the European Communities, OECD and the World Bank. The SNA is designed to give a realistic and compact view of the economy that is suitable for policy and analytical use.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: European System of Accounts.

Takeover (of enterprise)

Takeover of an enterprise occurs when its activities are taken over by another enterprise. As a result, the number of existing enterprises is reduced, and one of the original enterprises does survive in a recognizable form, therefore there is both continuity and survival. If two enterprises integrate entirely, one of the enterprises may remain largely the same. In this case the other enterprise is generally much smaller, it is merely absorbed by the larger enterprise, which remains the same. The takeover of enterprises is not considered to be real death. This event can be seen as the opposite of a split-off.

Source: European Commission, Eurostat, "Business Registers Recommendations Manual", Methodologies and Working papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: Cessation (of business), death (of enterprise), merger.

Takeover (of enterprise group)

Takeover of an enterprise group likewise describes the process where groups of enterprises integrate, the number of existing groups is reduced and, of the original groups, one keeps its identity. Enterprise groups may have many kinds of intergroup relations and integrate their operations partly or totally. Two (or more) enterprise groups may integrate entirely and become one group. In this process either both groups involved may lose their identity, because they are dissolved beyond recognition in the new

organization, or one group may remain largely the same. In the latter case the other group is generally much smaller; it is merely absorbed by the larger group, which remains largely the same. If one of them keeps its identity, it is called a takeover.

Source: European Commission (Eurostat), OECD, *Eurostat – OECD Manual on Business Demography Statistics*, Methodologies and Working papers, Publication Office of the European Communities, Luxembourg, 2008.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf

Related terms: Cessation (of business), death (of enterprise group), merger.

Truncated enterprise group

A truncated enterprise group consists of the enterprises and the legal units that are resident in the same country and belonging to the same multinational enterprise group. It may comprise only one unit, if the other units are non-resident. An enterprise may be the truncated group or part thereof. A truncated group may consist of several units and subgroups, which may appear unlinked if their parent is non-resident, but they actually belong to the same multinational enterprise.

Source: European Commission, Eurostat, *Business Registers Recommendations Manual*, Eurostat Methodologies and Working Papers, Publication Office of the European Union, Luxembourg, 2010.

Link: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf

Related terms: All-resident enterprise group, enterprise group, multinational enterprise group.

Turnover

Throughout this publication, the terms "turnover", "sales" and "revenues" are used interchangeably.

Source: United Nations, *International Recommendations for Industrial Statistics 2008*, para. 4.144 and table IV.1.

Synonyms: Revenue, sales.

Ultimate controlling institutional unit

The ultimate controlling institutional unit of a foreign affiliate is the institutional unit, proceeding up the chain of control of a foreign affiliate, where that chain is not controlled by another institutional unit.

Source: Eurostat, "Foreign AffiliaTes Statistics (FATS) Recommendations Manual", Methodologies and Working papers, 2012.

Link: http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-12-016

Related terms: Institutional unit.

Unincorporated enterprise

An unincorporated enterprise represents the production activity of a government unit, NPISH or household that cannot be treated as the production activity of a quasi-corporation. An unincorporated enterprise is a producer unit which is not incorporated as a legal unit separate from the owner (household, government or foreign resident); the fixed and other assets used in unincorporated enterprises do not belong to the

enterprises but to their owners, the enterprises as such cannot engage in transactions with other economic units nor can they enter into contractual relationships with other units nor incur liabilities on their own behalf; in addition, their owners are personally liable, without limit, for any debts or obligations incurred in the course of production.

Source: 2008 SNA.

Link: http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Related terms: Enterprise.

Value added tax (VAT)

VAT is a tax on products collected in stages by enterprises. Producers are required to charge certain percentage rates of VAT on the goods or services that they sell. VAT is shown separately on the sellers' invoices so that purchasers know the amounts that they have paid. Producers, however, are not required to pay to the government the full amounts of the VAT invoiced to their customers because they are usually permitted to deduct the VAT that they themselves have paid on goods and services purchased for their own intermediate consumption, resale or gross fixed capital formation. Producers are obliged to pay only the difference between the VAT on their sales and the VAT on their purchases for intermediate consumption or capital formation, hence the expression value added tax.

Source: 2008 SNA.

Link: https://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Variable

Variable is synonymous with the term "characteristic" but is more appropriate in the case of surveys where sampling is involved and thus the notion of variability due the probability mechanism applied for selecting samples.

Related terms: Characteristic.

Vertical integration

The vertical integration of activities is given wherever the different stages of production are carried out in succession within a unit and the output of one process serves as input to the next.

Source: United Nations, Statistics Division, "International Standard Industrial Classification of all Economic Activities (ISIC)", Statistical Papers Series M No. 4, Rev. 4, New York, 2008.

Link: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf

Related terms: Activity, horizontal integration.

Bibliography

African Development Bank. *Guidelines for Building Statistical Business Registers in Africa*. Tunis, 2014. Available at www.afdb.org/en/knowledge/publications/guidelines-for-building-statistical-business-registers-in-africa/.

Asian Development Bank. *User Guide for ADB Statistical Business Register*. Manila, 2018. Available at www.adb.org/publications/adb-statistical-business-register-user-guide.

Barnabé, Richard. Seeing the whole elephant: A proposed experiment on measuring the activities of multinational enterprises, *Statistical Journal of the United Nations Economic Commission for Europe*, vol. 20, No. 2 (January 2003), pp. 173–183. Available at www.unece.org/fileadmin/DAM/stats/documents/ces/2003/13.e.pdf.

Bérard, Hélène, Stuart Pursey and Eric Rancourt. Re-thinking Statistics Canada's Business Register, Ottawa: Statistics Canada. 2005. Available at https://nces.ed.gov/FCSM/pdf/2005FCSM_Rancourt_Berard_Pursey_IVB.pdf.

Council of the European Communities. Seventh Council Directive 83/349/EEC of 13 June 1983 based on Article 54(3)(g) of the Treaty on consolidated accounts. Available at europa.eu/legislation_summaries/internal_market/businesses/company_law/l26010_en.htm.

_____Council Regulation (EEC) No. 696/93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community. Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1993R0696:20081211:EN:PDF.

Daas, Piet, and others. Checklist for the quality evaluation of administrative data sources. Discussion paper 09042, CBS. The Hague, 2009. Available at www.cbs.nl/nr/rdonlyres/0dbc2574-cdae-4a6d-a68a-88458cf05fb2/0/200942x10pub.pdf.

Daas, P., and others, "Big data as a source for official statistics", *Journal of Official Statistics*, vol. 31, No. 2 (2015), p. 255.

Departamento Administrativo Nacional de Estadística. *Metodología Directorio Estadístico de Empresas – DEST,* Bogotá, D. C., Colombia. Versión 5. 2013. Available at www.dane.gov.co/files/investigaciones/fichas/Met_DEST_13.pdf.

ECB, OECD and Eurostat, Final Report by the Task Force on Head Offices, Holding Companies and Special Purpose Entities (SPEs), Paris, 2013, p. 8.

European Union Commission Regulation (EU) No. 446/2014 of 2 May 2014 amending Regulation (EC) No. 295/2008. Available at https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32014R0446.

_____ Regulation (EC) No. 177/2008 of the European Parliament and of the Council of 20 February 2008 establishing a common framework for business registers for statistical purposes and repealing Council Regulation (EEC) No. 2186/93 (Business Register Regulation 2008). Available at http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32008R0177.

_____ Regulation (EC) No. 295/2008 of the European Parliament and of the Council of 11 March 2008 concerning structural business statistics. Available at https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32008R0295.

_____ Regulation (EU) No. 549/2013 of the European Parliament and of the Council of 21 May 2013 on the European system of national and regional accounts in the European Union. Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:174:0001:0727:EN:PDF.

_____ Regulation (EU) 2019/2152 of the European Parliament and of the Council of 27 November 2019 on European business statistics, repealing 10 legal acts in the field of business statistics. (European Business Statistics Regulation 2019). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R2152&from=EN.

Eurostat. *Handbook on Data Quality Assessment Methods and Tools.* Wiesbaden, 2007. Available at http://ec.europa.eu/eurostat/ramon/statmanuals/files/Handbook_on_data_qual_assess_tools.pdf.

_____ *Business Registers Recommendations Manual.* Eurostat, Methodologies and Working Papers, Luxembourg: Publications Office of the European Union, 2010. Available at http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf.

_____ *Methodology for Manual European Profiling.* 2014. Available at https://ec.europa.eu/eurostat/cros/system/files/essnet-esbrs1_del-wp3-a1.1_methodological-report-profiling.pdf.

_____ *Handbook on Methodology of Modern Business Statistics.* 2017.

_____ *Handbook on the Compilation of Statistics on Illegal Economic Activities in National Accounts and Balance of Payments.* Manuals and Guidelines, Luxembourg: Publications Office of the European Union, 2018. Available at https://unstats.un.org/unsd/nationalaccount/docs/KS-05-17-202-EN-N.pdf.

_____ *European business profiling: Recommendations manual, 2020 edition.* Manuals and Guidelines, Luxembourg: Publications Office of the European Union. 2020. Available at https://ec.europa.eu/eurostat/documents/3859598/10479728/KS-GQ-20-002-EN-N.pdf.

Eurostat, *European Business Statistics Manual*, 2021 edition. Available at https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-21-001.

Eurostat and OECD. *Eurostat-OECD Manual on Business Demography Statistics.* Paris: OECD Publishing, 2008. Available at https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-ra-07-010.

Fellegi, Ivan, and Alan Sunter. A theory for record linkage, *Journal of the American Statistical Association*, vol. 64, issue 328 (1969), pp. 1183–1210.

Food and Agriculture Organization of the United Nations. *Conducting Agricultural Censuses and Surveys.* FAO Statistical Development Series, No. 6. Rome: FAO, 1996. Available at www.fao.org/fileadmin/templates/ess/documents/publications_studies/publications/SDS_6_Conducting_agricultural_censuses_and_surveys.pdf.

Herzog, Thomas, Fritz Scheuren and William Winkler. *Data Quality and Record Linkage Techniques.* New York: Springer, 2007.

ILO, Recommendation concerning the Transition from the Informal to the Formal Economy, 2015 (No. 204), para. 37.

_____ Resolution concerning statistics of employment in the informal sector, adopted by the Fifteenth International Conference of Labour Statisticians on 28 January 1993. Available at www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_087484.pdf.

_____ *Women and Men in the Informal Economy: A Statistical Picture.* Third edition (Geneva, International Labour Office, 2018).

_____ *Measuring Informality: A Statistical Manual on the Informal Sector and Informal Employment.* Geneva: ILO, 2013. Available at www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_222979.pdf.

IMF, *Balance of Payments and International Investment Position Manual*, sixth edition (BPM6) (Washington, D.C., 2009), paras. 4.54–4.55.

International Organization for Standardization, ISO 9000: 2005 para. 3.1.1.

Jaro, Matthew. UNIMATCH: a computer system for generalized record linkage under conditions of uncertainty. In *Proceedings of the May 16-18, 1972, Spring Joint Computer Conference.* New York: Association for Computing Machinery, pp. 523–530, 1972.

Joint Economic Commission for Europe/Eurostat/OECD Group of Experts on Business Registers. Papers and proceedings of meetings. Available at www.unece.org/stats.

Karr, Alan, Ashish Sanil and David Banks. Data quality: a statistical perspective. National Institute of Statistical Sciences, Technical Report No. 151 (March 2005).

Lane, Marc J., *Social Enterprise: Empowering Mission-Driven Entrepreneurs* (Chicago, Illinois: American Bar Association, 2011).

OECD, *OECD Benchmark Definition of Foreign Direct Investment*, fourth edition (Paris, 2008), p. 232.

OECD and others. *Measuring the Non-Observed Economy: A Handbook.* Paris, 2002. Available at www.oecd.org/std/na/1963116.pdf.

Ritzen, Jean. Statistical business register: content, place and role in economic statistics. CBS, Introductory Overview Lecture presented at the Third International Conference on Establishment Surveys, session II. Montreal, June 2007. Available at http://ec.europa.eu/eurostat/ramon/coded_files/RITZEN_Jean_SBR.pdf.

Skalitz, Anne. The French Business Register: From a quality approach to a statistical register, INSEE France, paper presented at the 21st Meeting of the Wiesbaden Group on Business Registers. Paris, November 2008. Available at www.oecd.org/industry/business-stats/41691270.pdf.

South African Statistical Quality Assessment Framework, second edition. Available at www.statssa.gov.za/standardization/SASQAF_Edition_2.pdf.

United Nations. *International Standard Industrial Classification of All Economic Activities. Revision 4.* Statistical Papers Series M No. 4/Rev. 4, New York, 2008. Available at http://unstats.un.org/unsd/publication/seriesM/seriesm_4rev4e.pdf.

_____ *Accounting for Global Value Chains: GVC Satellite Accounts and Integrated Business Statistics.* New York, 2019. Available at https://unstats.un.org/unsd/business-stat/GVC/Accounting_for_GVC_web.pdf.

_____ International Recommendations for Industrial Statistics 2008 New York, 2009.

_____ *United Nations National Quality Assurance Frameworks Manual for Official Statistics.* New York, 2019. Available at https://unstats.un.org/unsd/methodology/dataquality/.

_____ *Satellite Account on Non-profit and Related Institutions and Volunteer Work*, ST/ESA/STAT/SER.F/91/Rev.1 (New York, 2018), para. 1.6.

United Nations and others. *System of National Accounts 2008*. New York, 2009. Available at http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf.

_____ *System of Environmental-Economic Accounting 2012 Central Framework.* New York, 2014. Available at http://unstats.un.org/unsd/envaccounting/seeaRev/SEEA_CF_Final_en.pdf.

United Nations Committee of Experts on Business and Trade Statistics. Papers and proceedings of meetings. Available at https://unstats.un.org/unsd/business-stat/UNCEBTS.cshtml.

United Nations Economic Commission for Europe. *Generic Statistical Business Process Model,* Version 5.0, December 2013. Available at www.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0.

_____ *Guidance on Modernizing Statistical Legislation* (Geneva, 2018). Available at https://unece.org/statistics/publications/guidance-modernizing-statistical-legislation.

_____ *Guidelines on Statistical Business Registers.* New York and Geneva, 2015. Available at https://unece.org/info/publications/pub/21867.

_____ *Guidelines on the Use of Statistical Business Registers for Business Demography and Entrepreneurship Statistics.* New York and Geneva, 2018. Available at https://unece.org/sites/default/files/2021-09/Guidelines%20final.pdf.

United Nations and Eurostat. Working Paper on Statistical Business Registers, 2011.

Wiesbaden Group on Business Registers. Papers and proceedings of meetings. Available at https://unstats.un.org/unsd/methodology/citygroups/wiesbaden/.

Vennix, K., ICES IV, *The treatment of large enterprise groups within Statistics Netherlands*, 2012.

Winkler, William. String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 354–359, 1990.