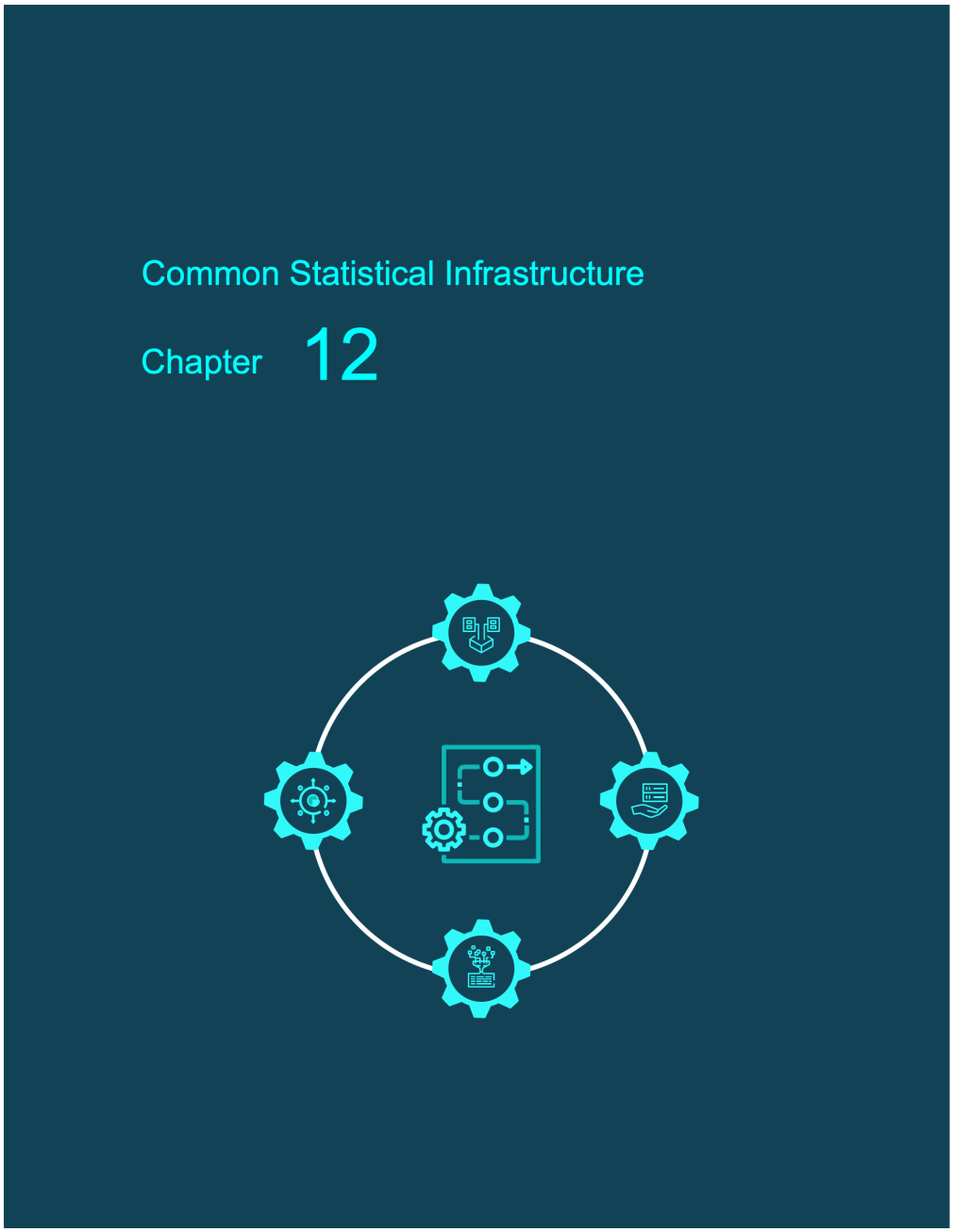
CHAPTER

TWELVE

# COMMON STATISTICAL INFRASTRUCTURE




## 12.1 Introduction

### 12.1.1 Definition of common statistical infrastructure

> The common statistical infrastructure is defined as the statistical tools and systems that support the activities constituting a generic statistical production process -but that is not part of any specific statistical process- together with the statistical activities required for development or acquisition, maintenance and promotion of these tools and systems. The common statistical infrastructure is part of the so-called intangible assets of a national statistical office (NSO) or any other producer of official statistics within the country.

Examples best illustrate the distinction between activities associated with a generic statistical process and with the common statistical infrastructure:

- As the first example, consider the subprocesses of designing and creating the frame for a survey as described in the Generic Statistical Business Process Model (GSBPM) subprocesses 2.4 and 4.1. In the case of a business survey, the frame is typically sourced from a statistical business register (SBR), which is not part of the survey, rather it is developed and maintained for the benefit of all business surveys. Thus, the SBR is an element of the common statistical infrastructure and associated with the statistical activities required to design, develop and maintain it and promote its use. Likewise, an address register is an element of the common statistical infrastructure providing survey frames for household surveys.
- As the second example, consider the subprocesses of designing and building a survey questionnaire (GSBPM subprocesses 2.3 and 3.1). They are typically supported by a questionnaire design and construction tool/system independent of the survey. This tool/system is another element of the common statistical infrastructure and is developed or purchased, maintained and promoted, for the benefit of many surveys.
- As the third example, consider the subprocess of checking output tables for confidentiality preservation (GSBPM subprocess 6.4). This is typically supported by a confidentiality checking tool, which is also independent of the survey, and which is developed or purchased, maintained and promoted for the benefit of many surveys. This tool is another element of the common statistical infrastructure.
- As the fourth example, consider the development, maintenance, and promotion of a statistical standard that defines all the ways a country can be acceptably divided by geographical area to collect and disseminate statistical data. Such a standard is used in the design and dissemination phases of a generic statistical process but is independent of the process. It is another element of the common statistical infrastructure.

### 12.1.2 Benefits of common statistical infrastructure

A common statistical infrastructure has two key benefits. First, each element of the infrastructure, once developed, supports all statistical processes, or, at least, all statistical processes of a given type, say business surveys. Thus, a new statistical process can be developed more quickly and at less cost than if there were no common statistical infrastructure, and it can be conducted more efficiently and with greater coherence with other processes. Second, the common statistical infrastructure promotes harmonisation across statistical processes through the use of common methods and standards.

Evidently, the NSO and other large producers of official statistics are likely to have considerably more common statistical infrastructure than smaller producers and, as such, they are likely to derive more benefits. However, smaller producers may also benefit if, as members of the national statistical system (NSS), they can take advantage of common statistical infrastructure developed by the larger members. In fact, the benefits of a common statistical infrastructure apply not only to producers in the NSS, but also to producers in different countries. This is the major aim of UNECE's Common Statistical Production Architecture (CSPA) (🔗), which is defined as:

> "a set of principles for increased interoperability within and between statistical organizations through the sharing of processes and components, to facilitate real collaboration opportunities, international decisions and investments and sharing of designs, knowledge and practices."

CSPA is further described in *Chapter 15.2.12 — Common Statistical Production Architecture (CSPA)*.

### 12.1.3 Content and structure of the chapter

*Chapter 7 - Users and their Needs*, *Chapter 9 - Data Sources, Collection and Processing*, *Chapter 10 - Analysis and Analytical Frameworks*, *Chapter 11 - Dissemination of Official Statistics* and *Chapter 8 - Quality Management* describe and provide guidance for the statistical activities associated with a generic statistical production process as described by the GSBPM. The current chapter is complementary to these chapters in the sense that it focuses on the statistical tools and systems that support the activities directly associated with a generic statistical production process while being separate from them. However, it does not cover statistical training, statistical information, data, and metadata management, which are dealt with in *Chapter 13 - Human Resources Management and Development* and *Chapter 14 - Data, Information and Knowledge Management*.

A statistical production process starts with a conceptual target population, which is defined in terms of the type of unit that is the subject of the process (for example, enterprise, person, household, farm, etc.) and the envisaged coverage of these units (for example, all enterprises registered for VAT, all persons that are permanent residents of a country or region, etc.).

As discussed in *Chapter 9 - Data Sources, Collection and Processing*, for a statistical production process that is a survey, the target population is realised in the form of a set of sampling units in a survey frame (abbreviated frame where the context is clear). In principle, the survey frame should provide complete unduplicated coverage of all units included in the target population. In practice, it is the closest approximation to the target population that can be reasonably obtained. The population covered is termed the survey population. The survey frame is not simply a list of sampling units; it includes all the data about those units required for stratification, sampling, and contact.

Survey frames should be coordinated across the surveys that are in any way related so that the data resulting from the surveys are coherent, i.e., can be readily combined without anomalies. For example, the frames for two business surveys intended to cover different industrial sectors should have mutually exclusive coverage.

How a survey frame is constructed depends upon the type of sampling unit involved. In this context, surveys may be divided into four basic types: business surveys, informal sector surveys, agricultural surveys and household surveys.

For most business surveys, the frame is most efficiently and effectively constructed by using a statistical business register (SBR). This describes the construction and maintenance of an SBR and its primary function - that of providing frames for business surveys - and its other uses, particularly as a source of publishable business demographics.

To the extent it depends upon administrative data for construction and maintenance, an SBR cannot provide coverage for informal sector surveys. Frames for such surveys can be obtained from an economic census based on area enumeration,

as a by-product of a population and housing census, or using a household-based survey in which a sample of individual households is asked about the businesses.

> Although business, informal sector, agricultural and household surveys are treated separately, the three groups are becoming increasingly interrelated.

For example, household surveys may be used to collect data about businesses in the informal sector or identify agricultural producers for own consumption. The SBR may be used as the frame for agricultural surveys of incorporated businesses that are in agriculture. Close cooperation between a population census and agricultural census in a country is strongly advocated in the relevant international guidelines.

## 12.2 Register-based statistics

### 12.2.1 Description of Register-based statistics

'Register-based statistics' mean statistics that are fully based on administrative registers or partly based on them complemented with data from other administrative sources and direct surveys. This definition is rather wide and may include almost all statistics production. The focus in this section is on those statistics which have links to a register-based production model.

The term 'register-based statistical system' is used in this section even though it is by nature indicative rather than an accurate description of any statistical system. As a pragmatic approach the handbook *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices* (🔗) published by the UNECE (2011) defines register-based statistical system as a system based primarily on administrative data that have been organized into linked statistical registers. This definition considers that in practice, all NSOs, independently of their data collection policies and production models, need to conduct regular statistical surveys to produce a full range of statistics. **However, in a register-based statistical system, the basic data infrastructure is organized around linked statistical registers based on administrative sources**. With such an infrastructure, administrative data is considered as a main source of information for the compilation of official statistics. Administrative data has been used in modern statistics production for decades and a remarkable milestone in producing register-based statistics was the Danish population and housing census of 1981 and its underlying register-based production model. The production model was based on the linking of several administrative registers and other administrative data sources. Since then, several NSOs have applied this approach to the production of register-based statistics.

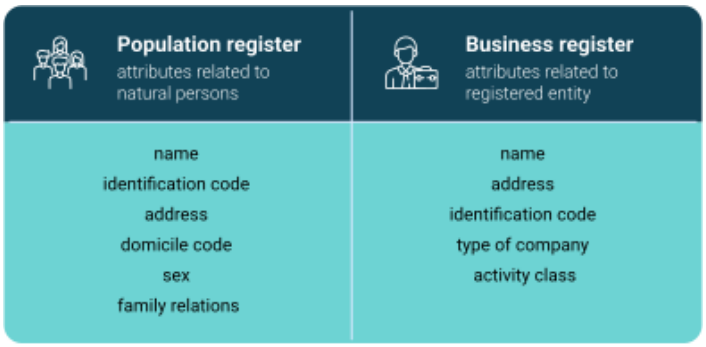
The above mentioned UNECE Handbook summarises several countries' experiences in collecting and using administrative data at that time. A common understanding among statisticians about the possibilities and constraints of using administrative data has grown remarkably during the recent decade. At the same time, statistical methods to improve possibilities and overcome problems have developed. It has become a good international practice to add a specific section to the use of administrative sources to domain-specific guidelines, such as guidelines for population census or integrated business statistics, to remind the NSOs about the possibilities of administrative data as a source for compiling statistics.

Administrative sources, access and acquisition of administrative data are dealt with in *Chapter 9.3 — Administrative sources*. This section deals with utilising administrative registers and other administrative data effectively in the statistical production. The focus is on building an administrative and statistical register infrastructure to be used in the production of official statistics - including technical and organizational structures and facilities - which can facilitate the use of register data.

> The utilisation of administrative sources varies greatly from country to country. The extent of the development of the basic infrastructure in the direction of intensive utilisation of administrative sources depends on the existence of suitable administrative sources, ease of access to them, general facilitators like common identifiers and, in the end, on government policy and the acceptance of the general public. The development phase is long, and a step-by-step approach is needed.

### 12.2.2 Registers

**Administrative registers**




A register can be simply a basic list of all units in the target population and nothing more. In practice, most registers include additional attributes for each unit. In a population register, these attributes are related to natural persons: name, identification code, address, domicile code, sex, and family relations. In a business register, a registered entity may have attributes like name, address, identification code, type of company and activity class.

An administrative register's function is typically to identify the registered units, keep stock of the population (like natural persons in a population register and businesses in a business register) and keep track of any information changes related to coverage (new units to be added and exiting units removed) and to events (changes in the attributes of the units). Management of these changes is built into the updating system of the register.

All countries have administrative registers. Administrative registers are seen increasingly often as a resource and facilitator in the effective and proper functioning of a society. Governments are giving ever-growing attention to the development and management of the administrative registers and the information management of public administration in full including information systems, information security, operating processes and technical interfaces. As part of the public sector information systems, administrative registers are needed to serve not only government administration in its daily operations but also to help people, businesses, and other organizations understand their lawful rights, access benefits, and meet their obligations in the society.

Among the most important administrative registers widely used for statistical purposes are population registers or civil registration and birth and death registration systems, business registers, registers of social security and health care systems, taxation and customs registers and registers from building and property registration systems.

The way the registers and other governmental information systems are organized and developed varies from country to country and depends on the culture, policies and structure of the governmental sector as well as on the legislative framework of a society. The degree of centralisation of administrative registers, the identification systems and the extent to which the registers and the identification systems are regulated by law vary remarkably across the countries.

Identifiers play a vital role in maintaining administrative registers and their use for statistics production, which usually means linking data from various sources. Identification codes (ID codes) should ideally not be changed during the period a unit exists. This is the method usually applied in the countries with a common identification system. The codes are used in administrative data across the whole administration.

> Registers and identification codes are usually regulated by national laws in the countries with centralised management of public sector's information systems. Laws regulate the responsible authorities and their rights and obligations as well as the content, access to and use of these registers. The creation and the use of the personal ID code and the business ID code are also regulated by law.

The national legislation on protecting private life and other basic rights, which safeguards an individual right to privacy, governs society. For these reasons, the number of authorities with access to the population register and the right to handle any information with personal ID codes is very limited. Within this legislative framework, the NSO has the right to get access to and use record-level administrative data are guaranteed, and provisions to safeguard statistical confidentiality are included. By contrast to the population registers and other administrative registers with private information, in some

countries, the business register's information about ID code, like address, place of residence and main activity are public by law and available on the web.

The legislation and organization of the registers, the updating systems, and the identification systems used in them, create the framework in which the national statistical system operates. Indeed, a centralised register-based administrative system with unique identification codes facilitates the acquisition of administrative data for the production of official statistics. When the administrative registers are more decentralised between government and administration entities, and especially if a common identification system used across the administration does not exist, an NSO has to make more effort to acquire data and process them. NSOs have made efforts for matching data from various administrative sources without a common identification system, but these efforts are very laborious, and the results may be quite weak in the context of regular statistical production.

*Chapter 9.3 — Administrative sources* discusses administrative sources and related quality issues in general. Due to the diversity of administrative register practice, data quality assessment frameworks for administrative data are bespoke. The use of administrative registers in statistical production requires that the NSO assess each administrative register's relevance, suitability, and quality as a data source. The coverage, data content and updating system of the administrative register should fit the statistical purposes and meet the overall data quality requirements. The updating system for an administrative register has a significant impact on its coverage and data content at a certain point in time. Therefore, it is vital for statistical usability that the register contains reliable time references for registrations and updates. Administrative registers and systems to manage them have to be stable over time to meet the continuity requirements for statistics. The data should be available in the form agreed between the producer of data and the NSO.

***Statistical registers and register-based approach to production***

> The NSOs keep statistical registers exclusively for statistical purposes. They are needed to organize the NSO's data collection activities by providing suitable frames for target populations and sampling. Statistical registers are usually created by processing data from several administrative sources, with or without combining them with survey data. Administrative registers are very seldom directly suitable as statistical registers, but they form an excellent source. A statistical register typically plays the role of a data coordination tool integrating data from several administrative and statistical sources. The data sources are integrated primarily by linking record-level data with common identifiers, occasionally by using matching techniques.

The generic model for register-based statistics is discussed in '*Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices* (🔗) published by UNECE. *Chapter 10.3 — Methods and systems of analysis* of this Handbook illustrates the links between administrative sources, statistical registers and statistical output.

The importance of statistical registers became obvious from the time when countries started to develop register-based censuses. Based on the available selection of administrative registers, the first step in the development process is to define a priority order in which the administrative registers play a role in the production system and which registers offer the possibility of identifying the units and links between the registers. This leads to selecting the most important ones to be used as a basis for statistical registers needed in the census system. Created statistical base registers, originally for census purposes, allow for implementing a meaningful system of internal links between these registers and the links to other registers and administrative data.

A statistical population register, a statistical register for buildings and dwellings or addresses and a statistical business register are usually defined as statistical base registers. Depending on the country's administrative registers and identification system, the statistical base registers may also include other statistical registers. The statistical base registers serve as a basis for the register-based system to statistical production. This is the main feature in the 'register-based production model'. The updating systems of the underlying administrative registers make it possible to choose the optimal updating dates for statistical base registers regularly updated at least annually. Suppose administrative records or other administrative registers than the main administrative source register are used in the statistical production; in that case, these additional data sets are linked to the corresponding statistical base registers. This 'centralized' register-based approach offers a basis for coordinating and integrating the production systems of population and social statistics based on administrative data, survey data or a combination of both.

The statistical business register cannot be based totally on administrative registers because it also contains information on establishments and enterprise groups, usually not available in administrative registers. The statistical business registers are dealt with more in detail in *Chapter 12.3 — Statistical Business Register*. Once the statistical business register with

an updating system has been created, it serves as one statistical base register in statistical production and is an essential element in integrating business and economic statistics. The register-based approach to production may decrease the number of direct statistical surveys, but it does not diminish their importance. Many data items and variables are not found in any administrative register or data source. These data items include, for example, details of the activities of people and businesses, and their opinions, expectations and behaviours. In all countries, there is always a great need for regular statistical surveys to produce a full range of official statistics.

Traditionally the statistical registers were created to serve as a target population for sampling frames. With the increased content and regular updating mechanisms, statistical registers are increasingly being sources of statistical data in their own right, particularly regarding population and business demography, and data for small areas or small populations subgroups. The role of statistical registers has increased especially since NSOs have become aware of, and started to develop, their common data architecture and data warehouses.

### 12.2.3 Production of register-based statistics

***Main features of register-based statistics***

The term 'register-based statistics' was first used in the 1980s in the context of register-based population census. At the same time, it was recognised that the same register-based approach used with established register-based production systems is also useful in producing many other statistics.

> The basic principle of the register-based production model is that all source data, from administrative sources and direct surveys, use the same identification system and link to the statistical base registers.

The most common way to produce statistics using the register-based system is to combine data from different administrative sources or combine data from administrative sources and direct surveys.

The register approach allows the production of some register-based statistics without any complementary data from direct surveys. These statistics may include population and business demographic statistics, education statistics, crime statistics and housing statistics. For instance, in the register-based production model, demographic variables are produced in the statistical population register and are used in all kinds of population and social statistics.

***Register-based population census***

The "*UN Principles and Recommendations for Population and Housing Censuses*" (🔗) published in 2017 offers useful information on census methodology summarising in its Chapter 4 the methodological questions related to the three data collection methods: full-field enumeration, register-based census, and combined methodologies, and concluding with the advantages and disadvantages in them.

A register-based population census requires a complex data system mainly based on the linking of various administrative registers. The combination of administrative registers and data used varies across countries as do the methods used and the production models applied. Development of a register-based census has usually taken many years, or even decades, depending on the development of suitable administrative registers and infrastructure needed in the system. The missing administrative data have usually been complemented with new or existing data from direct surveys during the development phase.

> Typically, a register-based population census requires record-level data from several registers with unified identification codes and an appropriate way to organize them into a register-based census system.

Population registers, registers for buildings and dwellings or addresses and business registers are usually the cornerstones in this system. These contain the links to other registers and provide links to other administrative sources. Depending on the national registration systems, other typical administrative registers used in register-based censuses are registers for taxation, employment, pensions, social welfare, job seekers and students. The register-based census system in four Nordic countries (Denmark, Finland, Norway and Sweden) is thoroughly explained in the publication "*Register-based Statistics in the Nordic Countries – Review of Best Practices with Focus on Population and Social Statistics* (🔗). The Dutch paper "*The usability of administrative data for register-based censuses*" (🔗) gives an illustrated description of the development process towards a fully register-based census in the Netherlands. It explains the country's registration system and offers a good description of the solutions and methods developed to overcome some of the most important issues.

Administrative registers do not usually cover all required data or provide enough details to allow the production of all statistical variables for the census. To complement the information that is not available from registers, countries use various methods that best suit their respective national circumstances. The methods used in ten different register-based population censuses are described in the document *Efficiency in Population Censuses – the situation of the European register-based 2011 Censuses* (🔗).

Once the register-based census system has been created, the infrastructure often allows annual production of the main statistical data contained therein. The register-based population census system usually offers a good basis to produce geospatial data in the countries with interlinked registers. One way to organize the production is via the register of buildings in which buildings and dwellings are geolocalised with map coordinates. Using the register-based inter-operability features, the exact location of each statistical unit can be derived. Thus, most variables of the population and housing census can be geo-localised.

The register-based census system also offers a good basis for developing special services for researchers. These services have currently developed rapidly, containing a large amount of anonymised and interlinked micro-data on persons, households, and businesses. These may be offered to researchers as online services. However, in giving access to census data on a more detailed level (microdata), it is crucial to secure confidentiality and privacy as discussed in *Chapter 10.3.1 — Methods of analysis*.

***Register based approach in the production of other statistics***

Vast amounts of statistics in many countries are based on a combination of data from administrative sources and data from direct surveys. These may typically include annual and short term economic and business statistics, income statistics and statistics on social conditions of households as well as a growing amount of energy and environment statistics. These statistics are linked in the register-based production model to statistical base registers so that the domain-specific statistical systems can use administrative data more easily and combine administrative data with survey data. The aim is to increase coherence and consistency in statistics production. The infrastructure developed for register-based statistics offers possibilities in fostering data warehouses and data integration in official statistics production. "*A Guide to Data Integration for Official Statistics*" (🔗), published by the UNECE, and the "*Asia-Pacific Guidelines to Data Integration for Official Statistics*", issued by ESCAP describe data integration in detail regarding different source data, methods and tools for data linking and matching.

Detailed administrative records can be an excellent source of information to compile more complex statistical frameworks, such as the national accounts or environmental-economic accounts.

> A register-based system with a vast amount of administrative and survey data stored and structured in a meaningful way in statistical registers and warehouses may allow for the production of complex statistics without launching new surveys. Further, this system increases the flexibility and agility to respond to new and emerging needs for statistics and indicators which is of growing importance in view of the monitoring the progress towards the SDGs. As for statistical surveys, statistical base registers function as sample frames.

A register-based statistical system also contributes to increasing sample surveys' quality, making them more consistent with other statistics at the macro-level and serving as additional information in analysing the results. The statistical base registers contain important demographic data for units that can be used to define populations and select samples. Using statistical base registers, samples may also be drawn for desired sub-groups, such as employed persons or students. In the analysis phase, survey results may be compared to data of the desired reference group available in the statistical base registers and other sources linked to them.

Statistical registers are also used to mitigate the impact of non-response in sample surveys when register data is used for the imputation of missing or invalid values.

### 12.2.4 Infrastructure for production of register-based statistics

***Coordination and working with data providers***

Production of register-based statistics usually requires a formal organization that supports the extensive co-operation between register authorities and an NSO. Ideally, the authority and coordination function of the NSO is recognized beyond the national statistical system, such as with administrative data holders, users, and other stakeholders. The organization model of coordination varies from country to country. A good practice has been that the NSO names a senior statistical expert to function as a coordinator for each of the most important producers of administrative registers and other data. Cooperation and coordination, as well as the follow-up, are often managed by a functional coordination unit at the NSO, which is also acting as an information hub between the different producers. The conclusion of formal written agreements (or Memoranda of Understanding/MoU) between the NSO and administrative data holders belongs to such a functional coordination unit's essential task in a register-based statistical system. These contracts help all parties to understand their obligations. Some authorities with many administrative data systems, like tax authority, may prefer centralized coordination mechanisms which are helpful also to the NSO. The designated coordinator at the NSO and his/her counterpart or contact person in the administration prepare the MoU for the data delivery or access with detailed technical attachments for each separate data set to be signed at the top management level. Attachments which contain specific detailed information of each data set including technical details and delivery timetables need to be updated regularly. The expert(s) role in the coordination unit is to manage and update the overall contract system and liaise between the subject matter specialists in the NSO and the holders of administrative data.

> NSOs operating in a register-based statistical system also need to actively participate in the development of public sector information systems to safeguard the basis and continuity of the data sources for the production of statistics and eventually find new possibilities for further development.

Therefore, it is a good practice to consult and invite the NSO to participate in any governmental initiatives that may have an impact on the accessibility, inter-operability, scope, content, periodicity, and timeliness of administrative data.

***Other aspects of common infrastructure***

Within a register-based statistical system, the general infrastructure of an NSO needs to be adapted to the growing flows of incoming administrative data complemented by survey results. There is often a need to re-engineer and integrate the processes and metadata systems, develop methods and quality frameworks, streamline the internal organization and policies, and develop staff capabilities. **Complementing or substituting survey data with administrative data for the production of official statistics without adapting the statistical production process rarely works**. At a general level, the statistical process does not change, but the use of administrative data in a register-based system impacts every phase of the process.

Wide methodological expertise is needed to utilise administrative data in statistics production, such as data matching and other methods used to integrate administrative sources into the production process. Increasing amounts of administrative data used and linked in the statistical process increase the need for proper disclosure methods and rules. **It is also important that the quality frameworks, rules and practices are adapted to this production system**. The staff working with register-based statistics may need additional competences and training. Those experts working with administrative registers and other administrative data in the production of domain-specific statistics need to know how to handle both, survey data and the administrative data they use. The register-based production system should be reflected in human resource policies and programs. A special function for administrative data, often as part of the data collection unit, may be useful. Its task would be to make the first validation and quality checks for incoming administrative data, review the production processes and acquire or develop software applications. This unit, as mentioned above, would also be responsible for managing the ToRs.

As discussed in *Chapter 14 - Data, Information and Knowledge Management* production of statistics needs well-functioning, uniform and standardised metadata systems. In register-based systems, it is important that administrative registers and other administrative data used in production be documented in the metadata systems. Careful documentation of administrative source data, record descriptions, and electronic access to questionnaires, including instructions should be stored in the metadata system. Also, all changes should be documented and stored. Processing rules and a possibility of tracing data to the source are an essential part of this documentation. The NSOs have developed over the years common databases for metadata used in the whole production process. These may include metadata for statistical

units, for classifications with various levels, concepts, variables and characteristics with definitions as well as metadata for technical standards. This kind of standardised metadata repositories improves the consistency of statistics and streamline the production process. These metadata repositories should be planned and adapted to the multi-source production process.

- *Efficiency in Population Censuses – the situation of the European register-based 2011 Censuses* (🔗);
- *A Guide to Data Integration for Official Statistics UNECE* (🔗);
- *Using Administrative and Secondary Sources for Official Statistics. A Handbook of Principles and Practices. UNECE 2011* (🔗);
- *Principles and Recommendations for Population and Housing Censuses. Revision 3, UN 2017* (🔗);
- *The usability of administrative data for register-based censuses* (🔗);
- *Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics. UNECE 2007* (🔗).

## 12.3 Statistical Business Register

### 12.3.1 What is an SBR?

> Narrowly defined, a **statistical business register (SBR)** is a set of businesses, including all the characteristics of those businesses required to provide frames for business surveys. More broadly interpreted, an SBR is these data together with the procedures, systems and human resources that support their use. In operational terms, it is a regularly updated, structured database that contains organizational, administrative and statistical units describing businesses in a country or region, that is maintained by a dedicated staff, supported by a purpose-built software system and that is used by an NSO to create frames for business surveys and sometimes as a direct source of counts of businesses by type, size, geography, economic activity, etc.

*Chapter 12.2.1 — Description of Register-based statistics* provides an overview of the key elements of a well-designed SBR, detailed in the subsequent subsections. Indications of how current SBRs may fall short of this standard are also provided.

***Primary role and benefits of an SBR***

The primary role of an SBR is as the source of frames for business surveys. In this context, to be of good quality, the SBR should contain all the businesses relevant to the surveys that use it together with accurate and up to date information about these businesses required for them to be stratified, samples and contacted.

The reasons why having an SBR as the common source of business survey frames is better than for each individual survey to create and maintain its own frame are that:

- an SBR facilitates the application of common standards and classifications, for example the International Standard Industrial Classification of All Economic Activities Revision 4 (ISIC Rev 4) across all business surveys;
- an SBR harmonises the surveys to which it provides frames in terms of coverage and units, thus facilitates the integration of the resulting business statistics;
- an SBR is effective; it provides well defined and better coverage than individually constructed survey frames, are likely to do; and
- an SBR is efficient; it ensures that data on key characteristics (such as main economic activity, location, turnover, and number of employees) are collected once only and shared across surveys; this saves resources and allows survey staff to focus on collection and analysis rather than frame preparation.

Sometimes the SBR is used simply as an input for frame development by a survey that maintains and augments its own local copy of the SBR for its own purposes. This is certainly not ideal practice but is done where the facilities for updating the (central) SBR are poor or difficult to control.

***Creation and maintenance of an SBR***

An NSO can construct the sets of units within an SBR in essentially two different ways: first, by field enumeration, typically during the course of an economic census; and second by using registers created for administrative purposes by other government departments and agencies. Maintenance of these units is by use of data from administrative sources and data from statistical sources.

***Economic census as the basis of an SBR***

Use of an economic census as the basis for an SBR was once a common practice in all NSOs, and even now it is used in a number of countries. For example, according to the *AfDB's SBR Guidelines* (🔗), Botswana, Mozambique, Uganda, and Zambia have all developed an SBR from an economic census within the last decade. However, the approach is no longer used by countries with well-developed SBRs for the following reasons.

- First, an economic census is very expensive even if its scope is restricted to businesses with identifiable premises. The numbers of businesses and costs are even further increased if businesses operating from households without visibility are also covered. Thus, an economic census can be conducted at best every five or ten years. Thus, well-developed NSOs consider a suite of annual surveys based on a register-based list to be much more cost-effective.
- Second, an economic census produces a point in time list of businesses. The small businesses located during area enumeration are volatile in the sense that they may go rapidly in and out of production, or ownership, or change their activities or addresses. Thus, to be really effective as a source of survey frames over the years until the next census, the list has to be constantly updated.
- Third, the only means of updating the list is by a full-scale ongoing enumeration operation. NSOs cannot afford to do this, so the list becomes steadily more out of date, resulting in survey samples containing inactive or untraceable units.

Nevertheless, in some countries, there is a case for conducting a periodic economic census and/or area enumeration in order to conduct surveys of the informal sector.

***Administrative input data***

> **Administrative data** are data collected by an **administrative source** (meaning a government department or other agency) **primarily for administrative (not research or statistical) purposes**, such as taxes, annual permits, etc. Administrative data should provide most of the data about the small businesses contained in the SBR and a starting point for obtaining and maintaining data about large businesses. The reason for using administrative data is that they provide comprehensive coverage of registered businesses, they are constantly updated, and they are essentially free to an NSO.

In a country with good access to administrative data, collectively, these data determine the coverage of the SBR and hence of the survey frames it provides. In countries where access to administrative data is limited, statistical data may have to be used to provide adequate coverage.

The quality of administrative data is an important consideration.

> Close coordination between an NSO and the administrative source is imperative to ensure the administrative data are well understood and that there are no surprise changes in the data flow or content.

Legislation, a memorandum of understanding, a service level agreement, or similar arrangement, and the use of a common identification coding system, facilitate data flow and data sharing.

***Statistical input data***

Statistical input data refer to **data collected by statistical processes undertaken by an NSO**. They are used to complement the administrative input data about businesses in an SBR by providing extra information about:

- the organizational structure, and appropriate reporting arrangements, for very large complex businesses, which are obtained by profiling, i.e., research and discussion with those businesses; and
- characteristics of businesses that are not available from the administrative data or are not current; these characteristics may include economic activity code, turnover, employment or other size code, current operation status (active, inactive, out of business), and contact details.

***Informal sector***

Businesses that are not registered with any of the sources used to create the SBR are typically small. In the national accounts, they are referred to as household unincorporated enterprises. Those that produce market output may be defined to constitute the informal sector. Information about them cannot be obtained from surveys based on the SBR. This is an important limitation in the scope of the SBR, and of the surveys based on it, especially in developing countries where the informal sector's contribution is relatively large.

> Data for the informal sector can be obtained by economic census based on area enumeration, or as a by-product of a population and housing census, or using a household-based survey in which a sample of individual households is asked about the businesses that (i) they conduct and (ii) they have not registered with any of the administrative sources used to construct the SBR.

This is further discussed in *Chapter 12.4 — Frames for informal sector surveys*.

***SBR Outputs***

The primary output of the SBR is a periodic (typically quarterly) common frame which is a snapshot of the active units in the SBR as of a given date, from which individual survey frames are extracted and business statistics generated.

In addition to providing survey frames, the SBR has other purposes.

First, it is a potential source of business demographic statistics; Second, it facilitates measurement of the response burden imposed on businesses by surveys; and Third, it provides a framework within which unit-level data from different surveys and administrative sources can be brought together.

Furthermore, the SBR system can readily be extended to include sample selection (from the survey frame) and generation of the corresponding sample control file used in subsequent conduct of a survey. Such an extension ensures a smooth interface between these functions.

***SBR Processes***

Figure 10 summarises the SBR inputs, processes and outputs, and their interactions in a fully developed SBR. The central block labelled SBR Database is often referred to as the live register. The data it contains is continually being updated based on data from administrative and statistical sources.

Administrative data sources may include tax registration data, VAT transaction data, income tax data, employer pay-as-you-earn data, company registration data and license data. To avoid the risk of duplication, it is best to use only sources having a common identification scheme or known to have mutually exclusive coverage. Updating should be fully automated, with only small numbers of incoming data records flagged for manual verification to ensure that SBR staff are not overwhelmed.

Statistical data are obtained as feedback from sample surveys and censuses that have derived their frames from the SBR and from SBR operations, including profiling of large businesses, SBR quality improvement surveys of smaller business and ad hoc investigations. Again, to the extent possible, such input is automated.

Periodically (typically quarterly) a snapshot, is taken of the SBR database. It provides a stable basis for analysis as the live register continues to evolve. It is sometimes referred to as a frozen frame. Numbers of units by industry, size code region, and numbers or units for which characteristics are missing, are derived and used in planning future SBR operations.

Periodically (typically quarterly or annually) a common frame, is derived from a snapshot, including only those units that are marked as being active and that include sufficient information to be stratified for sample selection and contacted if selected. The common frame provides the basis from which survey frames are extracted. It may also provide the basis for the publication of business statistics.

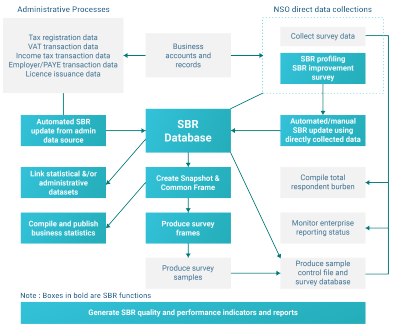



Fig. 12.1: Typical SBR inputs, functions and outputs

A sample is selected from each survey frame, a sample control file and embryonic survey database are created, and the survey conducted with data being feedback to the SBR, as noted above. Based on the survey control files, response burden can be monitored.

The SBR Database may also be used as a means of linking other administrative and statistical datasets.

***SBR Units Model***

Figure 11 indicates a typical SBR units model, i.e., a view of the types of units used in describing the SBR, its inputs and its outputs. The units are in three groups:

- **organizational units** – comprising legal units that are businesses (companies, partnerships or sole proprietors), groups of legal units defined by ownership and control arrangements, and operational units into which (large) businesses organize themselves, such as divisions, branches, plants, and outlets;
- **administrative units** – such as income tax unit, value-added tax (VAT) unit and employer/pay-as-you-earn (PAYE)) unit; these are the units that businesses use in responding to the various administrative requirements; usually (but not always) there is one unit, or no unit, of each type associated with a business; and
- **standard statistical units** – comprising (in this particular model) a three-tier hierarchy (enterprise group, enterprise and establishment) that the NSO uses to represent businesses in a standardised way for sample selection and data collection purposes. Some NSOs may use a different and possibly more complex model.

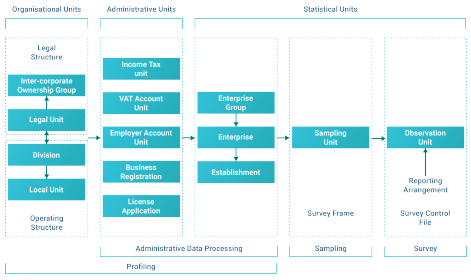



Fig. 12.2: Typical SBR Units Model

### 12.3.2 Conceptual framework for SBR

As described in *Chapter 10.4.1 — System of National Accounts (SNA)*, 2008 SNA is the core element of the conceptual framework and most important harmonizing and integrating mechanism for economic statistics. It defines the notion of economic production activities in which institutional units in their role as enterprises engage. The definition of an enterprise is very broad. It includes commercial enterprises, (comprising corporations and unincorporated household businesses) and it includes non-commercial enterprises (comprising non-profit institutions and government units). It includes enterprises in the informal sector, and even households producing goods for their own consumption (which are not included in the informal sector).

The 2008 SNA and the *International Standard Industrial Classification of All Economic Activities, Revision 4* (ISIC Rev 4) note that an enterprise, particularly a large enterprise, may be engaged in a range of different activities at a set of different locations. In such a case, the classification of a large enterprise to a single activity at a single location will result in a blurring of detail. This will constrain subsequent analysis of any datasets that include data collected from the enterprise. Thus, it is recommended that large enterprises are divided into smaller, more homogeneous producing units that can be more precisely classified and that collectively represent the enterprise as a whole. Dividing an enterprise into its various economic activities at its various locations (in a process referred to as profiling and further described below) results in a set of establishments, each of which refers to a particular activity at a particular location. An enterprise may alternatively, or as well, be divided simply by its economic activities into kind of activity units (KAUs) or by its geographic locations into local units.

The 2008 SNA also refers to conglomerates of enterprises – enterprises that belong to an intercorporate ownership group. The behaviour of an enterprise may depend on its role in a group and this needs to be taken into account in defining the survey reporting arrangements. (For example, though part of the ownership structure, an enterprise may be a "shell company" with no production activities and hence out of scope for a production survey.) Thus, where appropriate, enterprises are associated with an enterprise group. The process of dividing enterprises into establishments and/or grouping them into enterprise groups is commonly referred to (though not in the SNA) as profiling.

In summary, **the SNA provides the basis for defining the set of standard statistical units by which all businesses large or small can be modelled**. Figure 11 indicates a particular three-tier hierarchical model (enterprise group, enterprise and establishment) that is in accordance with the 2008 SNA and is often used in practice. A common alternative is a three-tier model containing enterprise group, enterprise, and kind of activity unit. More complex models can be defined based on the 2008 SNA, for example, including kind of activity unit, local unit and establishment, but are rarely used in practice.

### 12.3.3 Administrative data inputs for SBR construction and coverage

As previously outlined and indicated in Figure 10, a well-developed SBR is created and updated based on administrative data and statistical data.

The first step in constructing an SBR is to select the **primary administrative data source**. This should be the source that most nearly provides the target SBR coverage and content. Figure 11 indicates the typical options, for example, business income tax data, VAT data, and business registration data.

As no single source is likely to provide all the coverage and content provided, data from the primary source are typically supplemented by data from secondary administrative sources. However, although there are a wide variety of possible administrative sources, combining them can be difficult. There is a serious risk of duplication unless the sources share a common identification numbering system, or they are known to be mutually exclusive. Thus, the introduction of a common identification numbering system should be promoted. In its absence use of record linkage techniques based on name, address and other characteristics to identify potential duplicates has been tried but rarely proved satisfactory. Thus, an SBR is typically based on a small number of carefully selected administrative sources that can be used in combination without risk of duplication. Jointly these sources determine the coverage of the SBR.

It is convenient and perfectly in accordance with the framework for the definition of the informal sector in Resolution II of the Fifteenth International Conference of Labor Statisticians 1993 (ICLS93) (🔗) to define the resulting coverage of

the SBR to be the formal sector and all other enterprises producing goods or services for sale or barter to be the informal sector.

Assuming that administrative sources provide adequate coverage, field observation by NSO staff should not be used to further expand coverage, first, because it is an inefficient use of resources, but mainly because the additional enterprises found by this method are likely to be transient and mobile and hence not worth trying to maintain in the SBR. Furthermore, the extra coverage that is obtained is difficult to quantify, i.e., it is difficult to explain precisely what coverage is still missing.

> While the use of administrative data may be impractical in countries where the NSO is unable to access such data or where they are of very poor quality, the aim should always be to maximise use of administrative data as they are constantly updated, and they are essentially free to the NSO.

### 12.3.4 Statistical sources to supplement SBR content

Although administrative sources provide the enterprises that define SBR coverage, they are not sufficient to maintain an SBR. They are typically deficient in one or both of two basic ways:

1. They do not provide the information required to divide large complex enterprises into establishments;
2. They do not provide current values of all the enterprise characteristics and contact information that are required to produce up to date survey frames. Characteristics required include activity status (active, inactive, defunct, etc.), economic activity code, and size measures (turnover, employment, assets, etc.). Contact data include name, physical address and geocodes, postal address, e-mail address, contact person(s), etc. Whilst administrative sources may provide current values of many characteristics; they may not provide values for all, or the values provided may not be current.

***Profiling***

As previously noted, to collect data from large enterprises efficiently and at an appropriate level of detail, large enterprises need to be divided into establishments and/or grouped into enterprise groups through a process commonly known as profiling. The first step in profiling is a desk study to identify, as precisely as possible, all the legal units belonging to an intercorporate ownership group and their inter-relationships. This is the basis for developing a first draft of the organizational structure and how it might be best represented in terms of the SBR units model (typically enterprise group, enterprise and establishment) for the purpose of collecting data. The second step is face-to-face contact with senior staff (typically the chief accountant or secretary) in the leading enterprise in the group to review and update the draft organizational structure and reporting arrangements. The final step is to create the corresponding statistical units in the SBR.

Evidently, profiling requires very high-level contact with the enterprises involved and is resource-intensive. Therefore, it is only worth doing for very large complex enterprises for which the identification of appropriate reporting arrangements is absolutely vital. Thus, only a very small number of enterprise groups are defined, and only a small number of enterprises are divided into multiple establishments.

***Feedback from business surveys***

To supplement and update what is received from administrative sources, current economic activity code, size measures and contact information are obtained as feedback from the units responding to business surveys that have drawn their frames from the SBR. Typically, the first page of a business survey questionnaire asks for verification or updating of contact information, and other parts of the questionnaire provide updated size measures and perhaps indications of changes in economic activity.

***SBR Quality Improvement Survey***

Survey feedback alone is insufficient to fully maintain the SBR as not all SBR units are included in surveys. Indeed, some are excluded precisely because there is insufficient information about them to determine if they are in scope for a survey, to sample them if they are, or to contact them if they are selected in the sample. To obtain this information, through its field staff, the NSO typically conducts ongoing special-purpose SBR surveys, referred to as nature of business surveys or SBR quality improvement surveys.

### 12.3.5 SBR snapshots and common frames

***Live register and snapshots***

As previously noted, the central block in Figure 10, labelled SBR Database, is often referred to as the live register, as the data it contains are continually being updated based on data from the various administrative and statistical input sources. It is possible to extract survey frames directly from this. However, as the database is constantly changing, many NSOs prefer to take a regular point in time snapshots of the SBR and use these as the basis for developing survey frames. Snapshots may also be used for SBR backup purposes, for analysis of SBR counts and content quality, and for direct production of statistics, as elaborated in the following paragraphs.

A snapshot of the entire SBR (live register) is the starting point for deriving SBR counts such as:

- numbers of enterprises by activity status (active, temporarily inactive permanently inactive, defunct, etc.); and
- numbers of enterprises, by region and by economic activity;
- data quality measures such as:
  - number of enterprises with no activity code;
  - number of enterprises with no turnover size code; and • number of enterprises with insufficient contact data to be able to locate the business.

***Common frame***

A common frame, also referred to as a frozen frame, is a subset of a full SBR snapshot obtained by selecting only those enterprises:

- that are active during, or at the end of, a specified reference period; and
- that contain sufficient information for the enterprise to be identified for inclusion/exclusion in the frame for any business survey being conducted for the given reference period.

A common frame is the starting point for extracting the frame for each survey conducted for the specified reference period. Use of this single dataset as the basis for defining the coverage of all surveys ensures consistency of the corresponding survey frames.

### 12.3.6 Generation of survey frames and samples

***Survey frames***

> The frame for an economic survey should be a subset of the common frame, comprising the set of statistical 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.

For example, the frame specification for a sub-annual manufacturing production survey may require all establishments that are active, have a manufacturing ISIC code, and have five or more employees.

***Sample Selection and Sample Overlap Control***

Given that the SBR is the source of survey frames, it is vital that the SBR frame extraction software interfaces nicely with the sample stratification and selection programs and initializes the sample control files and the input databases in which the survey data will be stored. Incorporating these functions within the SBR environment ensures they are well harmonised.

From the survey frame, a sample can be selected in accordance with the sample specification. For example, the sample specification for the above sub-annual manufacturing production survey may require that the sample includes all establishments with one hundred or more employees and a five per cent sample of all remaining establishments in each ISIC group. (Sampling is further discussed in *Chapter 9.2.4 — Survey design*).

The sample selection mechanism may include a requirement for the sample (to the extent possible) to avoid an overlap with the sample for another business survey or to have a specified overlap with that survey or with the previous cycle of the survey itself. This is best handled within the SBR environment.

### 12.3.7 Producing statistics directly from the SBR

The SBR is a source of business demographic statistics, i.e., counts of enterprises and of their births, deaths and changes by industry, size institutional sector, region, etc. Subject to quality considerations, these can be published.

Usually, the counts are based on the common frame so that they refer to a specific point in time and they do not include businesses that are not known to be active or for which key characteristics are missing. The aspects of quality that need to be considered are:

- the numbers of enterprise not included in the counts because due to missing data items in the SBR (for example ISIC code) that are not included in the common frame; and
- period to period changes in counts due simply to changes in SBR operations (for example cleaning up a backlog of enterprise with no activity code) and not to changes in the economy.

### 12.3.8 Other functions of the SBR

***Data linkage across administrative and statistical data sources***

The SBR is ideally placed to facilitate linkage of data at individual unit level across administrative and statistical data sources (subject to confidentiality provisions). For example, if business income tax is a data source for the SBR, then it is easy to link data from income tax returns (and the accompanying financial statements) to the data for the same business obtained via a production or employment survey.

***Respondent Burden Monitoring***

It is useful to be able to list all the surveys to which any given enterprise has to report within any given year and the status of those reports. It is also informative to measure the respondent burden imposed on enterprises by individual surveys and by all surveys in total. These functions are most efficiently handled through the SBR.

### 12.3.9 Sources of further information

***Guidelines on Statistical Business Registers, 2020 UNSD***

The *UNSD Guidelines on Statistical Business Registers* (🔗) were originally published by the UNECE in 2015 and subsequently revised by the United Nations Committee of Experts on Business and Trade Statistics in collaboration with the United Nations Statistics Division. As of June 2020, a white cover version prior to official editing is available.

The main objectives of the Guidelines are to:

- provide practical guidance on core issues of establishing and maintaining an SBR;
- clarify typology, concepts and definitions, including for statistical units;
- provide guidance on the use of administrative and other sources for the establishment and updating of an SBR;
- provide guidance on how to use an SBR in its own right for the 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; and
- provide guidance on the role of an SBR in the modernisation of statistical production and services.

The Guidelines are consistent with the 2008 SNA and its European version. They are targeted at SBR management and staff members, business survey staff, and staff dealing with the administrative authorities that deliver data to the SBR. They provide guidance in the form of a broad range of concepts and explanations that need to be interpreted within each particular national context.

***Guidelines for Building Statistical Registers in Africa, 2014, AfDB***

The design, development, and introduction or enhancement of an SBR is so important that it typically constitutes a core goal in the five-year national strategy of most if not all African NSOs. Thus, in response to requests from NSOs, the African Development Bank (AfDB) prepared the *Guidelines for Building Statistical Registers in Africa* (🔗). They detail all the essential elements of an SBR. Although produced with African countries in mind, the Guidelines are equally applicable to developing countries on other continents. They are consistent with the UNECE SBR Guidelines as they were input to the drafting of the latter.

The general objectives of the Guidelines are to provide:

- a general background to the need for an SBR and the concepts on which it is based;
- a description of the functions of an SBR;
- detailed information on SBR development and implementation; and
- a starting point for harmonization of SBRs across African NSOs.

The Guidelines are directed, first and foremost, at economic and labour statistics practitioners in NSOs and policy analysts in development, economic, and labour ministries and central banks. In addition, they may be useful to a wide spectrum of other users, including:

- **SBR managers and staff –** by detailing SBR concepts and creation and maintenance procedures, and SBR quality and performance measures;
- **business survey managers and staff –** by providing the basic concepts on which an SBR is based and describing the production of survey frames from the SBR;
- **business statistics dissemination staff –** by providing the basic concepts, including what data are included in the SBR and how these data may be published;
- **staff responsible for communications with other organizations in the national statistical system and with international organizations –** by providing the basic concepts;
- **staff responsible for respondent relations –** by defining and enabling calculation of the individual and cumulative respondent burden associated with business surveys; and
- **NSO senior managers –** by providing the basic concepts, quality and performance measures, and suggestions for quality improvement.

***European business statistics methodological manual for statistical business registers — 2021 edition, Eurostat***

The 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, sets out a common framework for the European business statistics, including the Framework of the European Statistical Business Registers.

The Manual aims to explain the reasoning behind the provisions of the Regulation as regard the Framework of European Statistical Business Registers, which includes the National Statistical Business Registers, the EuroGroups Register and the exchange of data between them (*European business statistics methodological manual for statistical business registers — 2021 edition - Products Manuals and Guidelines - Eurostat* 🔗). It aims to provide the extra information required for the correct and consistent interpretation and implementation in all countries and can be used as a tool to identify and recommend best practices.

***Statistics Canada Business Register***

The Statistics Canada Business Register (BR) is Statistics Canada's continuously maintained central repository of baseline information on businesses and institutions operating in Canada. As a statistical register, it provides listings of units and related attributes required for survey sampling frames, data integration, stratification and business demographic statistics.

The BR provides the statistical sampling frames for the 200+ business survey programs administered by the organization. It also serves as a central hub for data integration in the compilation of cross-cutting statistics obtained through record linkage. Finally, it is used to compile business demographic indicators, including the detailed counts of enterprises within industries and provinces that comprise the Canadian Business Counts product (further described below), which is released semi-annually.

- **Statistics Canada Business Register - Population**

  The BR maintains a complete, up-to-date and unduplicated list of all businesses in Canada that have a corporate income tax account, an employer payroll deduction remittance account, a Goods and Services Tax account, a partnership account, or a registered charities account. Persons reporting any of the various types of business income on personal tax forms are also included.

  The complex portion of the BR represents approximately 1% of the total active businesses on the database and accounts for approximately 52% of the total economic activity in Canada. The simple portion represents approximately 99% of the total active businesses on the database and accounts for approximately 48% of the total economic activity in Canada. The focus of manual intervention is on the complex portion, while automatic updates from administrative sources are made on the simple portion.

- **Statistics Canada Business Register - Data sources**

  The BR is updated by:

  - data from Canada Revenue Agency (CRA);
  - profiling of large and medium sized businesses (in the complex portion);
  - survey feedback - changes and corrections to BR frame data are transmitted regularly by survey collection areas during the collection of economic survey data; and
  - research tools such as the internet, provincial gazettes, trade and business publications and newspaper clippings.

- **Statistics Canada Business Register - Error detection**

  Editing of BR records is an on-going process that is performed daily by staff in various divisions with different roles across Statistics Canada. The Administrative Data Division takes in the CRA data and performs consistency checks and pre-processing activities. The Statistical Registers and Geography Division (SRGD), as the main manager and maintainer of the BR, then uses the data to build the register. Enterprise profilers and statistical officers on-staff in SRGD perform on-going verifications and updating of data. Industrial and other subject-matter economists working in the many survey program divisions also contribute to the maintenance of the frame. Given the significant number of editors, editing of BR data is controlled through an internal interactive system that ensures coherence and proper routing of editing tasks.

- **Statistics Canada Business Register - Disclosure control**

  Statistics Canada is prohibited by law from releasing any data including data based on the BR that would divulge information obtained under the Statistics Act that relates to any identifiable person, business or organization without the prior knowledge or the consent in writing of that person, business or organization. Some confidential data can be released for statistical or research purposes with the authorization of the chief statistician. Only government bodies can receive authorization from the chief statistician, permitting them to receive confidential data that identifies individual units (company names, addresses, etc.). Private companies are not permitted to receive this kind of data.

- **Statistics Canada Business Register - Demographic data**

Canadian Business Counts (🔗) provides counts compiled from the BR of active business locations according to variables, such as geography, business activity and employment size. It is not advised to use this product for time-series analysis involving comparisons across reference periods because of data accuracy considerations.

- **Coverage error**. The BR is largely based on the Business Number (BN) that is collected and assigned by Canada Revenue Agency. Therefore, the quality of the data is dependent upon the quality of the information submitted by Canadian businesses when applying for their BNs.
- **Under-coverage**. The BR is subject to a fluctuating number of unclassified BN records, and outstanding work and unassigned workloads within Statistics Canada.

***Australian Bureau of Statistics Business Register***

The Australian Bureau of Statistics (ABS) Business Register (BR) is a list of organizations that undertake economic activity in Australia. The data on the ABS BR is primarily sourced from the Australian Business Register (ABR), maintained by the Australian Tax Office (ATO), and via ABS profiling of large and/or complex businesses.

The scope of the ABS BR is all organizations with an active Australian Business Number (ABN) on the ABR, undertaking productive activity in Australia's economic territory. The relatively insignificant economic activity of organizations that fall below the threshold for needing to register for an ABN (i.e., annual turnover of less than $75,000 or 150{,}000$ for non-profit organizations) and that have chosen not to apply for an ABN, remain outside the conceptual scope of the ABS BR.

The characteristics and structural relationships of organizations on the ABS BR are described using the ABS Economic Units Model. The model defines organizations by enterprise group, type of activity, location and legal unit. It enables organizations to provide data to the ABS at suitable levels of detail according to their size and complexity.

The ABS BR provides a frame for most ABS economic surveys, thus enabling a consistent, coherent, point-in-time picture of the Australian economy. Data are extracted from the ABS BR on a quarterly basis, producing the Common Frame. Subsequently, survey frames are extracted from the Common Frame. The unit record information on the ABS BR is not available for release outside of ABS.

- **How the ABS BR is maintained**

  The ABS BR is a dynamic database that is updated daily via profiling activity and updated monthly using ABR and other ATO data.

  The ABS profiles large, complex and economically significant organizations and structures them in accordance with the model. These groups are collectively referred to as the Profiled population. The remainder of ABN registrants is assumed to have a simple structure comprising a single legal entity represented by a single enterprise group. These units are collectively known as the Non-profiled population. The two populations are mutually exclusive and cover all organizations in Australia that have registered for an ABN.

- **ABS BR - main outputs**

  The main outputs derived from the ABS BR are:

  - survey frames for use in ABS economic data collection;
  - information to support ABS economic survey processes including data editing;
  - a dataset that is the main input to the annual publication Counts of Australian Businesses, including Entries and Exits (🔗) (cat. no. 8165.0). This publication provides data relevant to users interested in understanding businesses that actively trade in goods or services.
  - aggregate business structure reports for internal analysis and research and to meet client requests by Australian and state government agencies and industry associations at the aggregate level; and
  - infrastructure to support the use of administrative-economic data.

***Philippine Statistics Authority Statistical Business Register***

The Philippine Statistical Authority (PSA) maintains a List of Establishments (LE), where the establishment is defined:

as an economic unit which engages, under single ownership or control, that is, under a single entity in one or predominantly one kind of economic activity at a single fixed physical location.

A Frozen LE is usually generated over the January to February period. From it survey frames are extracted, each according to its particular scope and coverage. Frames are provided to Annual Survey of Philippine Business and Industry, the Quarterly Survey of Philippine Business and Industry, the Labor Turnover Survey, the Occupational Wages Survey, the Integrated Survey on Labor and Employment, the Survey on Information and Communication Technology, and other establishment-based surveys.

An Annual LE is generated every June 30 and is the basis for the tables of establishment data disseminated to users.

The PSA updates it LE through the conduct of its Updating of the List of Establishments (ULE) operation. The ULE is a continuing activity undertaken primarily to provide an updated and reliable frame for the establishment- and enterprise-based surveys and census.

The updating activity involves:

- capturing "new" establishments and determination of their characteristics for inclusion in the LE;
- updating the characteristics of those establishments already listed in the LE; and
- tagging establishment records with the proper code reflecting their operational (activity) status as of the time of updating.

A comprehensive ULE is a nationwide undertaking which involves door-to-door canvassing and conduct of inquiry on the basic characteristics of establishments in the country. It is undertaken during the census reference year, that is, a year prior to the conduct of the Census of Philippine Business and Industry. A tablet-based inquiry form was utilized for the first time in 2018 Comprehensive ULE (🔗).

During intercensal years ULE operations cover only selected or growth areas, group of units with common characteristics, or group of industries with special concerns. The coverage depends mainly on the availability of resources and what is needed. Updating is usually supplemented by information from establishment's reports to surveys undertaken by the PSA, including the Census of Agriculture and Fisheries, which the SBR supports.

The LE is also updated from administrative data derived from the Central Bank, the Bureau of Internal Revenue, the Bureau of Customs, the Philippine Economic Zone Authority, and Securities and Exchange Commission. This is not all done on a regular basis; it depends upon what has been provided to PSA or can be found and downloaded by staff.

Profiling is undertaken to distinguish enterprises and establishments, and group of enterprises as well as to trace duplicate establishments. It is an effective method of identifying the units within an enterprise that are not yet captured in the LE. Ideally, profiling should involve input from the enterprise being profiled. Unfortunately, this does not happen. Instead, profilers use the information that they can gather from the enterprises' websites and other trusted sources.

***Business Register of the Statistical Institute of Jamaica (STATIN)***

**STATIN BR - Overview**

The Business Register (BR) of the Statistical Institute of Jamaica (STATIN) (🔗) is a structured list of establishments operating in Jamaica. It currently has approximately 12,300 establishments, organized by industry and assigned an activity code based on the 2005 Jamaica Industrial Classification (JIC). It is designed to:

- serve as a sampling frame for establishment surveys;
- consistently classify statistical reporting units, i.e., establishments;
- serve as a data source for compiling demographic information about businesses.

As of 2020 STATIN is working with Statistics Canada on the Project for the Regional Advancement of Statistics in the Caribbean (PRASC) (🔗) to improve the BR through review and redesign.

**STATIN BR - Coverage and Scope**

The BR covers privately operated businesses employing one or more persons. Establishments engaged in agriculture and entities in Central Government are not in scope. The BR covers both formal and informal businesses and contains the following variables:

- name of the establishment; year of commencement; type of organization;
- legal form of this establishment; type of ownership;
- business contact information; contact person information;
- economic activity; total number of employees.

**STATIN BR - Maintenance**

The BR is maintained through a programme of continual updating of business records. The process includes the following:

- Identification of business prospects (potential establishments for inclusion on the BR) from the internet, newspaper, telephone directory, etc;
- Administration of the Central Registry of Economic Units (CREU) questionnaire via a survey that is designed to provide updated information on businesses on the BR as well as initialize business prospects;
- Updating of records on the BR based on survey feedback (including CREU and other routine establishment surveys) and administrative records.

***Business Register (Tanzania Mainland) in the Tanzania National Bureau of Statistics***

- **TNBS BR - Overview**

  The Business Register (BR) of the Tanzania National Bureau of Statistics (TNBS) formerly known as the Central Register of Establishments (CRE) is a list of establishments operating in a single location with fixed premises. The variables included in the BR are:

  - name of the establishment; physical location (region, district, ward, street);
  - postal address including telephone; email and fax numbers; main industrial activity;
  - size group, ownership of establishment; registration status; number of people engaged by sex;
  - nationality of the owner of the establishment; turnover; and source of initial capital investment.

- **TNBS BR - Purpose**

  The BR provides a comprehensive inventory of establishments in Mainland Tanzania. It is used as a basis for generating or obtaining sampling frames for establishment-based surveys in the country.

- **TNBS BR - Population and Coverage**

  The BR covers all sectors of the economy and economic activities as stipulated in ISIC Rev 4. It includes all legal establishments, regional headquarters and rural and urban parts of some districts with a large number of businesses in all regions of Tanzania Mainland.

- **TNBS BR - Data Sources Used to Update the BR**

  BR updating is a continuous exercise done mainly by the regional statistical offices under the direct supervision of the Department of Field Operations. Data are obtained from the physical visits to new establishments and from administrative records from tax, license, pension fund and other registration authorities, including the Tanzania Revenue Authorities, Business Registrations and Licensing Agency, the Ministry of Health, Community Development, Gender, Elderly and Children, and the Ministry of Education and Vocational Training.

  The BR is also updated using data collected from economic surveys, including the Employment and Earnings Survey, Annual Survey of Industrial Production, and Tourism Statistics Survey.

# 12.4 Frames for informal sector surveys

## 12.4.1 Definition of the informal sector

The informal sector represents a very important part of the economy in many countries. SNA 2008 recommends the use of *Resolution II of the Fifteenth International Conference of Labor Statisticians (ICLS) 1993* (🔗) in defining the informal sector. The Resolution requires the informal sector to be defined as a subset of household unincorporated enterprises, comprising those enterprises that:

- produce at least some output for the market; and
- 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; and/or
- 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.

Enterprises that are involved in agricultural production may, optionally, be included or excluded; conceptually, they should be included.

The registration (third) criterion is becoming the most commonly used in practice.

It is important to recognize that these criteria were a practical compromise that provide a framework for a definition of the informal sector, rather than a single, unambiguous definition. They reflect variations across countries in how the informal sector is actually defined. For example, the Philippine Statistics Authority defines the informal sector as comprising business that are a single proprietorship in terms of legal organization, a single establishment in terms of economic organization and have a total employment of less than ten persons.

On the other hand, the *Guidelines for Building Statistical Business Registers in Africa* (🔗) suggest defining the informal sector as the set of non-agricultural household enterprises that have market production but that are not registered with the administrative source(s) on which the coverage of the SBR is based. This definition is in accordance with the Resolution and has the virtue of making the formal/informal boundary very visible and easy to implement. It means that:

- the suite of business surveys based on the SBR and administrative data measure the formal sector;
- non-agricultural household market enterprises not included in the SBR constitute the informal sector; and
- agricultural household market enterprises not included in the SBR constitute the informal agricultural sector.

At the 2018 International Conference of Labour Statistician (ICLS), the ILO received a strong mandate to initiate a revision of the 15th ICLS resolution concerning statistics of employment in the informal sector and the 17th ICLS guidelines concerning a statistical definition of informal employment. A working group has been established, and a new set of standards based on current definitions and country practices will be prepared for the next ICLS, presently scheduled for October 2023.

## 12.4.2 Relationship between the informal sector and SBR

As noted in the *Chapter 12.3 — Statistical Business Register*, a modern SBR is based on administrative data and provides coverage only of the businesses registered with the administrative sources, thereby excluding most or all (depending upon the definition) of the informal sector. The temptation to try and expand the coverage of the SBR to include the entire informal sector should be resisted as there is no sustainable way of doing so.

Thus, frames for surveys of the informal sector, or for an economic census that includes the informal sector, must be otherwise derived. This is addressed in the subsections below.

Three situations are discussed:

- frame for a survey of informal businesses with visible premises, based on area enumeration;

- frame for a survey of informal businesses conducted by households, based on a household frame, and
- frame for an economic census with an informal component.

The following subsections also indirectly address some of the problems in providing survey frames that are faced by countries that do not have a modern SBR either because there are no suitable administrative sources, or because the NSO cannot access them.

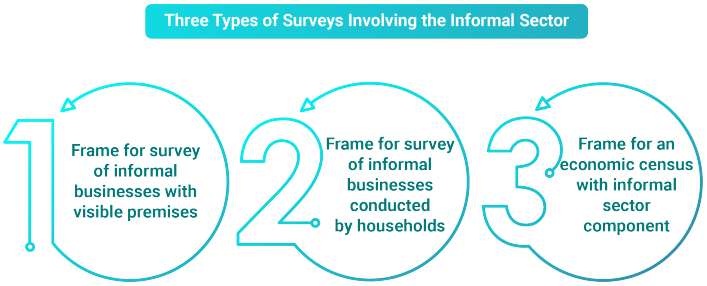
### 12.4.3 Measurement of the informal sector

The ILO publication *Measuring informality: A statistical manual on the informal sector and informal employment* (🔗) is a technical guide for NSOs. It focuses on technical issues that are specific to the production of statistics on informal employment and the informal sector. It presents the currenzt international standards on employment in the informal sector that were adopted by the ICLS in 1993 and the guidelines on informal employment adopted by the 17th ICLS in 2003. It provides practical guidance on implementing the standards and presents three basic approaches regarding the type of survey that can be used to collect the informal sector data, together with the strengths and limitations of each. They are:

- household-based surveys with a labour force component, notably labour force surveys;
- establishment-based surveys and censuses of production, and
- mixed surveys, where the first phase is a household survey, and the second phase is an establishment survey of the businesses obtained in the first phase; the first phase can be a standalone survey or the module of an existing household survey.

Although it describes these approaches in detail, the Manual should not be regarded as a comprehensive guide to labour force or other household surveys, or to enterprise surveys for which there are dedicated manuals. Also, it should be noted that a population census or household survey may collect data relevant to the informal sector (such as occupation, industry, class of worker, and place of work of household members) without specifically having an informal sector module.

The following paragraphs describe the provision of frames for three types of surveys involving the informal sector and the mixed survey approach.




***Frame for survey of informal businesses with visible premises***

As noted in the Manual, a comprehensive survey frame for informal businesses can be created only by area enumeration. It may benefit from information obtained by an earlier enumeration, for example from an economic census that has been conducted not too long previously. In any case, the country is divided into enumeration areas and a list of all businesses with visible premises within a sample of these areas is obtained and used as the basis for the survey frame. The set of businesses thus identified will not define the informal sector in accordance with the ICLS93 framework as businesses without visible premises are excluded.

As such a list is likely to fall rapidly out of date, unless the survey is to be repeated annually there is little or no point retaining the list for the next cycle. There is certainly no point in attempting to use it to create or update an SBR.

***Frame for survey of informal businesses conducted by households***

Such a frame can be created by area enumeration of households, or, more probably, by use of a household frame that has already been created for some other purpose, for example the conduct of a labour force survey or a population census.

Each household in the household frame is asked if they conduct a business and those that do form the frame for the informal sector survey of the businesses. During the course of the survey, those businesses that are found to be in the SBR are eliminated from the survey if the SBR is considered to define the formal sector.

> Such a frame gives broader coverage of the informal sector than a survey based on area enumeration of visible premises as it includes businesses conducted by households from their own homes without visible premises.

***Frame for an economic census with informal sector component***

The appropriate method for constructing a frame for an economic census depends upon the target coverage of the census. If the target coverage of the census is all businesses in the formal sector, then the SBR can provide the frame, and the results can be used to update the SBR. There is no coverage of the informal sector.

> If the target coverage of the economic census is all business with visible premises, then area enumeration as described above is required to cover the informal businesses with visible premises.

In a dual-frame approach, the resulting frame is used in combination with a frame from the SBR that covers businesses in the formal sector. To avoid duplication, businesses that are in the SBR frame and in the area-based frame, are identified, marked as duplicates and included only in the SBR frame.

Feedback from the census concerning businesses in the SBR is used to update the SBR. As the area-based frame cannot be maintained and is likely to fall rapidly out of date, there is no point in attempting to maintain it or use it to expand the coverage of the SBR.

If the target coverage of the economic census is all incorporated businesses and all household businesses, then a household-based frame as described above is required to cover the household businesses. It is used in combination with the SBR in a dual frame approach (as also described above). Again, feedback from the census concerning businesses in the SBR is used to update the SBR, but data about the informal sector business should not be used to expand the coverage of the SBR.

***Mixed surveys including the 1-2-3 Survey Approach***

The 1-2-3 survey is a type of mixed modular survey. It is an arrangement of three interlocking surveys, as shown in Figure 15. Sometimes only the first two phases are applied, referred to as a 1-2 survey.

The first phase is the creation of the frame by a household survey such as a labour force survey. Persons that are identified as owners and operators of informal household market enterprises form the frame for Phase 2. In Phase 2, more detailed information is collected regarding the business activities. Phase 2 can be carried out at the same time as Phase 1 or later (but not too much later). If later, a sub-sample can be used. Phase 3, if conducted, involves collecting data on consumption and living conditions.

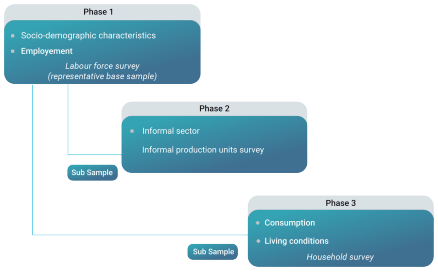



Fig. 12.3: 1-2-3 Survey

The 1-2-3 survey approach was first used in Mexico at the end of the 1980s. Initially designed to study the informal sector, the approach was gradually extended to measure and monitor poverty and governance. Over the last few years, the approach has spread to many countries in Africa, Asia and Latin America. The first two phases of a 1-2-3 survey are a labour force survey and an informal sector survey. The third phase is an income and expenditure survey, administered to a subsample of households identified in the first phase, and designed to estimate the weights of the formal and informal sectors in household consumption, by product and type of household. Ad hoc surveys can be added to any of the three phases to obtain additional information on, for example, access to micro-credit, social insurance and taxes.

### 12.4.4 Examples of national practices

***Afghanistan – Integrated business enterprise survey, 2009***

The integrated business enterprise survey carried out in Afghanistan was the first survey of its type in the country. It covered all non-agricultural economic units in all urban and most rural areas and accounted for about 81 per cent of the total rural population. As it lacked a complete up-to-date sample frame for all units in the country, it used a dual sample frame. A list-based frame was used for large establishments (generally with 20 or more workers) and an area frame for villages in rural areas and 'controller' (or enumeration) areas in urban areas. The area frame was constructed using the 2004 population census list of villages and 'controller' areas. The population censuses did not provide information on the concentration of establishments or their industries, so this information could not be used to create strata. Areas in urban and rural areas were therefore stratified only by province; in urban areas 'controller' areas were also stratified according to the concentration of establishments, with areas with markets or a high concentration of enterprises forming one stratum separate from other areas.

All large enterprises in the list-based frame were included in the sample.

The sample-based on the area frame was selected in two stages. In the first stage, 150 villages in rural areas were selected using systematic sampling with probabilities proportional to population. In urban areas, 150 controller areas were also selected with systematic sampling but with equal probability. Each controller area was divided into around five smaller segments, from which two were selected, one with the highest concentration of units and another at random.

In the second phase, all economic units within the selected villages and segments were listed. The listing included all units operating in fixed premises, with the exclusion of those already in the list-based frame, as well as all household-based and itinerant activities. This was achieved by inquiring in each dwelling, whether some economic activity was being carried out there or was based there. Activities carried out within the dwelling were included in the list, as well as activities of household members that were carried out in no fixed premises. Activities carried out by household members in fixed premises outside the dwelling were not included in the area frame, as they were listed at the village or controller area in which the activity was carried out.

***Ethiopia – Census of economic establishments, 2004***

The first census of economic establishments in Ethiopia was carried out by the Central Statistical Agency (CSA) in February-March of 2004. The census covered all urban establishments that were set up with the aim of making profits and covered both public and private establishments. The census included units that sold or produced in open markets or in a fixed location and mobile businesses in all sectors, with no limitation as to the number of persons engaged and the amount of income. It excluded civil service institutions and defence and other non-profit-making government and non-governmental organizations.

The fieldwork for the census was carried out without many problems. However, the reporting quality of some data items, particularly those on capital and revenue, was considered to be unreliable and the NSO was not capable of setting up a reliable SBR based on the data collected. Thus, the CSA's subsequent thrust is to establish a reliable SBR by conducting well-designed censuses of economic establishments and enterprises, and to use administrative data for supplementing and maintaining it.

***National survey of micro-enterprises (ENAMIN), Mexico***

Over the period 1992 to 2012 INEGI, Mexico conducted its *National survey of micro-enterprises* (ENAMIN) on a regular basis. The survey encompassed all non-farm activities without prejudging if they were formal or informal. For trade, transport, construction and services, it included units with six or fewer workers (including the owner or manager); for manufacturing, it included units with 16 or fewer (including the owner or manager). Mexico adopted only the two first stages of the 1-2-3 survey because an income and expenditure survey was already well established.

The first phase of the survey was the labour force survey, which has a questionnaire that already included the necessary elements for identifying entrepreneurs in the informal sector, as well as employees and contributing family members. In the second phase, the survey was directed to heads of non-farm micro-economic units (including professionals) operating either with or without premises, whether the business in question is the main or a secondary activity.

# 12.5 Statistical farm register and other frame sources for agricultural censuses and surveys

In the past, an economic statistic programme typically included a periodic economic census that was based on area enumeration of all establishments and that provided benchmark data and frames for production and other economic surveys conducted between censuses. As noted in *Chapter 12.3 — Statistical Business Register*, this situation has evolved so that fewer countries now conduct an economic census. Instead, data are obtained from administrative sources and from an expanded suite of economic surveys. The frames for these surveys are derived from an SBR, which itself is based on and updated from administrative data.

It could have been expected that agricultural censuses and surveys would evolve in much the same way, in other words that fewer countries would conduct an agricultural census, that data would be obtained from administrative sources and from an expanded suite of agricultural surveys, Eventually, the frames for these surveys would be derived from a statistical farm register (SFR) based on and updated from administrative data sources. However, whilst there may have been some movements in this direction, it has been inhibited by (at least) two significant constraints.

- First, there is a continued ongoing demand for data that can only be obtained from an agricultural census, or at least, a very large sample;
- Second, an SFR containing a comprehensive list of farms (or agricultural holdings, as the basic unit for agricultural census and production surveys is more precisely termed) cannot be constructed from administrative sources in the way that an SBR can. The administrative sources that exist are not sufficient. Therefore, inevitably, some agricultural frames have to be constructed, at least part, by area enumeration.

Although there are no international guidelines for SFRs, there are international guidelines for conducting an agricultural census that include how to construct the census frame, as further discussed in the *Chapter 12.5.2 — Frames for agricultural censuses*. There is also a global strategy to improve agricultural and rural statistics, which also has information about constructing the frames for agricultural surveys, as further discussed in *Chapter 12.5.3 — Methodology and guidelines for the generation of agricultural survey frames*. *Chapter 12.5.4 — Examples of National Practices* provides examples of national practices.

## 12.5.1 Sources of frames

The frames for agricultural censuses and surveys can be obtained by the following methods:

- full enumeration of all agricultural holdings in the country specifically for, and part of, an agricultural census;
- enumeration of agricultural holdings in a representative sample of areas specifically for an agricultural survey;
- full enumeration of agricultural holdings and own account agricultural activities during the course of a population and housing census;
- enumeration of agricultural holdings and own account agricultural activities in a representative sample of areas during the course of a household survey;

or from the following sources:

- cadastral or other land records that identify land parcels;
- an SFR that is created from data collected during an agricultural census and subsequently maintained by the NSO using any combination of the above enumeration methods and/or administrative sources and/or feedback from agricultural surveys;
- an agricultural master sampling frame (MSF), comprising a representative list of units, created from an agricultural census and/or a population and housing census, maintained by the NSO, and used as a general-purpose sampling frame for selection of samples for agricultural surveys;

- an SBR that is maintained by the NSO and that can provide a frame of businesses (in the formal sector) that are agricultural businesses.

In a situation where no single frame provides adequate coverage a dual frame approach may be used. Samples are selected from two frames, typically one being list based and the other area (enumeration) based. Taking care to avoid double counting of units that are in both frames, the results are combined to provide better overall coverage than could be obtained from either frame alone. The approach can be extended to three (or more) frames.

> In summary, although some countries do maintain a Statistical Farm Register (SFR), SFRs do not play the same major role in agricultural survey programmes as do Statistical Business Registers (SBRs) in economic statistics programmes. Indeed, as further elaborated below, many developed NSOs, such as the Australian Bureau of Statistics and Statistics Canada, no longer maintain a separate SFR. Instead, they may rely on a frame of agricultural businesses provided by the SBR coupled with an Agricultural Master Sampling Frame (MSF) of agricultural holdings obtained by enumeration in a representative sample of areas. This explains why there are no current internationally acknowledged guidelines, principles and recommendations for SFRs and why this section is entitled Statistical farm register and other frame sources for agricultural censuses and surveys rather than simply Statistical farm register.

### 12.5.2 Frames for agricultural censuses

***Scope of and approach to agricultural censuses***

As noted in the *World Programme for the Census of Agriculture (WCA) 2020, FAO* (🔗), the scope of an agricultural census varies depending on the resources available within a country, the importance of agriculture, the traditions in census organization, and national priorities. Thus, the census should be conceived as part of an **integrated system of agricultural statistics**. Typically, the main aim of an agricultural census is to provide basic data on the structure of the agricultural sector, such as size of holdings, land use, land tenure, etc. These are data that do not change quickly over time.

> Many countries do not have sufficient resources to conduct a full agricultural census, or a suite of agricultural surveys to replace it. Thus, they collect the data from a very large sample, still referred to as an agricultural census, or they collect only a subset of data items from all holdings and collect the remaining data from a sample.

The WCA 2020 recommends a modular approach to agricultural census and survey planning, with a core module based on a complete enumeration collecting a limited set of key items, coupled with sample-based complementary, thematic surveys. This approach is in line with the Global Strategy to Improve Agricultural and Rural Statistics (🔗) (further described in *Chapter 12.3.3 — Administrative data inputs for SBR construction and coverage*) and the *UN Guidelines on Population and Housing Censuses* (🔗). To facilitate implementation, the WCA 2020 provides guidance on agricultural censuses carried out by countries in the period between 2016 and 2025. It highlights the technical and operational benefits of coordinating the population census and the agricultural census. The *WCA Volume 1: Programme, concepts and definitions guidelines* (🔗), as further discussed below, are in two volumes in each of which there is practical information how to create frames.

***Statistical units of agricultural censuses***

The statistical unit for an agricultural census is an **agricultural holding** and is defined as 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, such as labour, farm buildings, machinery or draught animals.

There are two types of agricultural holdings:

- holdings in the household sector, that is, those operated by household members; and

- holdings in the non-household sector, such as corporations and government institutions.

For the non-household sector an agricultural holding often, but not always, coincides with an establishment (as defined for an SBR and business statistics), depending upon the exact definitions used for each.

In most developing countries, the majority of agricultural production is in the household sector. The concept of agricultural holding is therefore closely related to the concept of household. Standards for defining a household are laid down by the UN in its guidelines for population and housing censuses as follows:

"The concept of household is based on the arrangements made by persons, individually or in groups, for providing themselves with food or other essentials for living. A household may be either (a) a one-person household, that is to say, a person who makes provision for his or her own food or other essentials for living without combining with any other person to form part of a multi-person household, or (b) a multi-person household, that is to say, a group of two or more persons living together who make common provision for food or other essentials for living. The persons in the group may pool their resources and may have a common budget; they may be related or unrelated persons or constitute a combination of persons both related and unrelated".

For the household sector, there is usually a one-to-one correspondence between an agricultural holding and a **household with own-account agricultural production** activities (either for sale or for own use); in other words, all of the own-account agricultural production activities by members of a given household are usually undertaken under single management. There are two special cases where the concepts of agricultural holding and household with own account agricultural production may diverge.

- If there are two or more units making up a household, such as where a married couple lives in the same dwelling as their parents, the two units may operate land independently but, as members of the same household, they make common arrangements for food, and they pool incomes.
- In addition to an individual household's agricultural production activities, a household may operate land or keep livestock jointly with another household or group of households. In this case, there are two agricultural holding units associated with the household and two sets of activities: (i) the agricultural production activities of the individual household itself; and (ii) the joint agricultural operations with the other household(s).

Thus, some countries may prefer to define an agricultural holding in the household sector to coincide with a household with own-account agricultural production. The benefits of doing this are:

- identification of the holding in the agricultural census is simplified; it is no longer necessary to find out about multiple holdings within the same household;
- it is in line with the practice already used in previous agricultural censuses in many countries;
- the use of a common statistical unit – the household – enables the agricultural census to be more easily linked to the population census;
- it facilitates the analysis of household characteristics;
- if the scope of the agricultural census is expanded to include other households not engaged in own-account agricultural production, there is a common unit between agricultural production units and other households.

In defining the statistical unit for the non-household sector (corporations and government institutions), Volume 1 states that the concept of establishment (as an economic unit engaged in one main productive activity, operating in a single location) should be used.

***Agricultural Census Frame Construction***

- **Main types of agricultural census frames**

  There are two main types of frames for collecting agricultural statistics, a list frame and an area frame. A list frame comprises a list of agricultural holdings whereas the area frame comprises segments of land. In some cases, a multiple frame approach is used, in which part of the population (e.g., incorporated and government holdings) is covered by a list frame and the remainder (household holdings) by an area frame. For the census of agriculture, the list frame is the most common frame, and this is described below. An ideal frame is a list of all agricultural holdings, based on the operational definition of the agricultural holding adopted by the country, identifying each

unit without omission or duplication and without including any units other than agricultural holdings. Such a list could be obtained through a population and housing census for the household sector, an SFR, a listing exercise or other sources. The frame for the non-household sector can also come from an SFR, an SBR, administrative records or other data sources.

Whatever method is used, special care must be taken to ensure that all agricultural holdings are included in the frame. If holdings are missing, they cannot be enumerated in the agricultural census, and the validity of the census results is compromised. This is especially significant in an integrated agricultural statistical system, as any weaknesses in the agricultural census frame are also reflected in all the surveys that follow.

- **Agricultural census frame from SFR**

  Where an SFR exists, it is a good frame for an agricultural census, provided:

  – it includes all agricultural holdings according to the definition established; and

  – it is regularly updated, to remove units that cease to operate as holdings and to add new holdings.

  Along with the usual identification criteria, an SFR usually contains some basic information about each unit, such as land area, types of livestock kept, types of crops grown. This information is updated periodically. It is used for stratification when sampling techniques are used.

  An SFR can be created in various ways. It may be initially created at the time of an agricultural census and regularly updated thereafter, using information from various sources. In other cases, it may be established by law as part of an administrative process and updated annually. The problem with this type of register is that the information is required for administrative purposes and may not coincide with statistical needs. The unit in the register may not correspond with the definition of the agricultural holding. For example, the register may be based on cadastral or other land records in which each parcel of land is identified rather than the holding unit. A register-based on land ownership is not entirely suitable for an agricultural census because several people in a household may manage land separately. Thus, ownership and management are not always synonymous. Furthermore, the landowner is not the land operator if the land is rented out. Frames based on business registration or licensing procedures are not always suitable as they represent what the unit is licensed to do, not necessarily what it actually does.

  For the non-household sector, the SFR may be based on records from government regulatory agencies. Most countries have business registration and licensing systems. Membership information from industry associations may also be useful.

- **Agricultural census frame from population and housing census**

  Another type of frame covering the household sector of an agricultural census is one created from the population and housing census as a one-time exercise, but which is not kept up to date or maintained as an SFR afterwards. This is possible if the population and housing census include questions on agriculture that identify households involved in own-account agriculture. Although these are not the same as agricultural holdings, they provide a starting point. Based on the list an NSO can:

  – contact each household in the agricultural census;

  – ask each household about the household's own-account agricultural production activities and the management of agricultural activities in the household, to identify each agricultural holding; and

  – enumerate all agricultural holdings.

  The definition of the agricultural holding must be kept in mind. Where this is based on minimum size criteria such as land area, numbers of livestock, numbers of fruit trees, etc., these minimum size criteria need to be identified at some stage.

  For such a frame to be useful, the agricultural census needs to be undertaken as soon as possible after the population and housing census to ensure that the list of households involved in own-account agriculture is current.

- **Construction of an agricultural census frame from scratch**

If it is necessary to build a frame, the country is divided into suitable geographical units, referred to as enumeration areas (EAs), covering the whole in-scope national territory. Then each EA is visited, and interviews are conducted with local authorities or visits are made to households to identify all agricultural holdings in the EA. Population and housing censuses are usually conducted using EAs as the basic building block, and it is often possible for the agricultural census to piggyback on the population and housing census field system by using the same EAs and making use of maps and other field materials.

- **Multiple frame approach**

  Typically, a combination of frames is used for the agricultural census. Often, the household sector is enumerated based on the EA frame of the population and housing census, whereas a frame of agricultural holdings in the non-household sector is obtained from administrative sources. *WCA 2020 Volume 2: Operational Guidelines* (🔗), which elaborates the operational aspects of conducting an agricultural census, *Chapter 14 - Data, Information and Knowledge Management* provides detailed information on frame construction.

### 12.5.3 Methodology and guidelines for the generation of agricultural survey frames

***Global strategy to improve agricultural and rural statistics, 2010, FAO and World Bank***

The third pillar of the *Global strategy for improving agricultural and rural statistics* (GSARS) is integration of agriculture into national statistical systems. Integration is needed for several reasons. A common shortcoming in both industrialized and developing countries is that data are collected by sector, using different survey frames. In addition to the NSO, there may be several other NSOs involved in the collection and analysis of agricultural, fishery and forestry data, and there is insufficient coordination. Surveys may be conducted on an ad-hoc basis with no links to a common frame or the use of georeferenced units. Thus, it may be difficult to combine data from different surveys, for example, crop and livestock production, for in-depth analysis. Integration of statistical systems can help solve many of these problems.

A very important aspect of integration is the development of an agricultural master sampling frame (MSF) and its use in a coordinated data collection program, thereby ensuring good coverage and coherent statistics. An MSF provides the basis for the selection of probability-based samples of farms and households and enables the characteristics of farms to be connected with those of households, and with land cover and land use dimensions. The challenge in developing an MSF is that it must satisfy the needs of three statistical units: agricultural holding, household, and land parcel. In addition, these three units must be linked so that, for example, household income, health, and other factors may be compared to the farm's economic situation, and all of these to their general environmental impact. Where there is a one-to-one relationship between the three types of unit, economic, social, and environmental information can be collected from a single unit. The main difficulty in the development of MSFs occurs where there is not a one-to-one link between the agricultural holding and the household.

***Regulation (EU) 2018/1091, 2018 on integrated farm statistics***

The *EU regulation (EU) on integrated farm statistics* (🔗), which builds on the GSARS, provides direction and guidance to EU countries and is a source of guidance for others. It notes that, in order to avoid placing an unnecessary burden on agricultural holdings and national administrations, coverage thresholds should be established and that, for proper analysis of the structure of European agriculture, 98% of the utilised agricultural area and 98% of the livestock units on farms should be covered. For some Member States, these thresholds are too high for national purposes. However, the agricultural holdings below those thresholds are so small that a sample survey once per decade is sufficient to allow the estimation of their structure and the impact on production. The approach results in a reduction of cost and burden while still enabling the design of effective policy action to support and maintain small farming structures.

***Handbook on Master Sampling Frames for Agricultural Statistics***

In accordance with the second pillar of the GSARS, the *Handbook on Master Sampling Frames for Agricultural Statistics* (🔗) provides detailed guidelines and country examples for the development and construction of an MSF. The Handbook is intended as a reference document providing technical and operational guidance on various aspects of the development and use of an MSF in various country conditions, with an emphasis on developing countries. The Handbook recognizes the diversity of country situations and resources and consequently proposes various options.

***Content and construction of an MSF***

The population for an agricultural survey should be defined in terms of one of the three standard statistical units – agricultural holding, household or land parcel.

A comprehensive MSF is a listing of sampling units that, when associated with reporting units, provides complete coverage of the population of interest, as well as linking agricultural holdings to households and land dimensions.

An MSF may comprise:

- a list of the names of farm operators obtained through an agricultural census,
- a list of households derived from a population and housing census,
- a list of commercial agricultural enterprises that are not associated with a household, or
- a list of area units that are defined geographically; or
- combinations of these.

If the list is of farm operators, the reporting unit is the holding associated with the name, and the items of interest are the land that it operates and the crops and livestock on that land, the households associated with that land, and the geo-referenced land. The list may also include commercial agricultural enterprises that are not associated with households and/or households with livestock but no land. Households that provide agricultural labour but do not operate a farm are excluded, as are households with small plots for which production falls short of a given threshold. This sort of list provides the linkage between an agricultural holding and the household associated with it. However, it is not a complete MSF because it excludes rural households.

If the list is of households, the reporting unit is the agricultural holding, and the items of interest include the land to which it is associated, the crops and livestock on that land, and the geo-referenced land. If the list has been derived from a population census, it will include all rural households, as well as those that are not linked to land but that have livestock contribute to the agricultural labour force or are simply rural non-farm households. It will not provide statistical efficiency at the first stage of sampling if the number of households used as an indicator of size when selecting PSUs is not well correlated to the items of interest such as crop areas or livestock inventories. It can become a more useful MSF if a listing of commercial agricultural enterprises is included. Alternatively, the agricultural enterprises may be used as a separate sampling frame in the multiple frame sampling context.

If the list is of segments or parcels of land, the reporting unit can be the holding associated with the land or the household(s) associated with the land having items of interest (crops and livestock on the land). Rural non-farm households within the land parcel are also reporting units. Commercial agricultural enterprises are also reporting units. This list is an area sampling frame, as well as an MSF.

### 12.5.4 Examples of National Practices

***EU Farm structure survey***

The basic *EU Farm structure survey* (FSS) is carried out by all EU Member States using a common methodology every 3 or 4 years and once in ten years as a census. The basic unit is the agricultural holding. The coverage is all agricultural holdings which meet the minimum requirements set in the applicable legislation. The information collected covers land use, livestock numbers, rural development, management and farm labour input (including the age, gender and relationship to the holder of the agricultural holding).

The FSS provides comparable and representative statistics across countries and time. The data can be aggregated by different geographic levels (countries, regions, and for basic surveys also district level). They can also be arranged by size class, area status, the legal status of the holding, objective zone and farm type.

***Farm Management Survey, Statistics Canada, 2017***

The following description of the *Farm Management Survey, Statistics Canada, 2017* (FMS) indicates that Statistics Canada does not maintain a separate SFR. For surveys such as the FMS, Statistics Canada draws the survey frame from the SBR.

For other agricultural surveys, such as the Farm Produce Prices Survey (FPPS) (described in the following paragraphs) it draws the frame from the most recent Census of Agriculture.

The FMS is conducted every five years. It contributes to Agriculture and Agri-Food Canada's work on measuring management practices on Canadian farms. The information generated from this survey helps measure management practices in the Canadian agriculture industry, address federal and provincial policy needs and support the development of effective agricultural programs. This mandatory survey focuses on both livestock and crop operations, specifically the production of dairy, beef, poultry, pig, field crops, forage crops and vegetable, fruit, berry and nut crops.

The conceptual universe for the 2017 FMS was all active farms on Statistics Canada's Business Register. The following groups were excluded.

- Active farms with agricultural sales in 2016 of less than $10,000.
- Institutional farms (prisons, research stations, colleges).
- Farms located on Indian reserves.
- Farms for which more than 50% of the gross income in 2016 came from sales of greenhouse, sod and nursery products.
- Farms included on the 2017 Greenhouse, Sod and Nursery Survey frame.
- Farms without livestock inventory or crop area at the time of the 2016 Census of Agriculture.
- Farms located in the Yukon, the Northwest Territories and Nunavut.

***Statistics Canada Farm Product Prices Survey***

The *Statistics Canada Farm Product Prices Survey* (FPPS) is a monthly survey that produces provincial data on average prices received by farmers for agricultural products produced and sold. The primary purpose is to provide the pricing component of the Farm Cash Receipts series, which values the agriculture sectors' contribution to the Canadian economy. The prices are also used to calculate the Farm Product Price Index series. Agriculture and Agri-Food Canada and other federal departments use the data to develop, administer and evaluate agricultural policies and programs. Provincial departments and agencies use the prices for production and price analysis to support provincial stabilization programs.

The target population includes all Canadian agriculture operations as defined by the Census of Agriculture, as well as all marketing boards, agencies, commissions and federal and provincial government departments that collect data on producer prices or data from which prices can be calculated.

## 12.6 Household address register

### 12.6.1 Roles and benefits of address register

> A statistical household address register contains organized contact information for every household. It can be generated from census or administrative information, but it is most commonly generated by combining census with administrative address information.

Administrative address information, with official names of streets, is usually in the jurisdiction of the local government and only in rare cases is subject to central government intervention. Even though street/block names and numbers are administrative information, the degree to which they are standardised and encoded varies greatly between countries. Some countries do not have a database of street names, while other spatial units (including dwellings) are codified to the greatest detail and regularly updated.

Codifying spatial units generally depends on national practices and administrative traditions. Recent attempts for international standardisation have been made through the UN-GGIM by developing the *Integrated Geospatial Information Framework* (IGIF), aiming to bring a spatial component into every data layer of the government. The importance of spatial information and GIS systems for official statistics is emphasised and further elaborated in *Chapter 9.4 — Geospatial data*.

NSOs, particularly those in charge of census taking, have a direct interest in the quality of address information. Thus, they have (with different results) managed to influence administrative bodies to improve (and standardise) the address information systems. Key to maintaining a high-quality statistical address register is having access to administrative data that is regularly updated. Unfortunately, building registration is often a complicated administrative process that involves multiple institutions and involves different procedures (safety, administrative, cadastral, taxation, etc.) and may therefore be difficult to influence. Census preparation may provide an excellent opportunity to start the standardisation process between different institutions. Each standardisation usually requires many discussions, but the key to success is agreeing to use a common identification number. Presentation Building and dwelling register as a base for the production of geostatistical data (🔗) by Swiss Federal Statistics Office provides an illustrative overview of the path that has to be taken by an NSO that wishes to promote a standardised approach through government administration.

In most advanced countries an address register is geocoded, each entry has a unique ID and is a member of multiple spatial hierarchies (street & house number, census area, school district, court district, hospital jurisdiction, etc.). Further, in some countries, an address register contains information about every building and every dwelling (apartment) inside every building. In some countries (i.e., Ireland) unique postal codes are assigned to each dwelling. This is commonly referred to as **buildings and dwellings register**. Dwellings registers often contain other properties such as surface area, floor, number of rooms, utility access, number of occupants and others.

> Having a reliable address register improves survey data quality and reduces the data collection costs, as it eliminates the need for area enumeration/household listing (which is usually performed before Census or other major surveys).

Countries that do not have access to high-quality administrative information on dwellings and their occupancy have sought to supplement address information with additional sources. This approach is often referred to as **multi-source household address register** where administrative information on addresses (that may or may not include dwelling (apartment) numbers) is supplemented with information from utility companies, tax returns or other sources of information. Household address registers provide a tabular list frame that can be used for household surveys and population censuses, it reduces the costs of the Census as it eliminates the need for address canvassing and increases quality of the data through increased coverage, particularly if Census is conducted via the mail-in method.

In preparation for the 2020 Census round, some countries have started using Big Data approaches to gather information needed to update the household address register. Through agreements with post offices, they have started processing postal address information to update their address registers. Further, most offices perform in-office address canvassing using satellite imagery, thus significantly reducing the number of addresses that have to be canvassed in-field before the Census. An overview of the approach used by the Australian Bureau of Statistics to maintain and update Address Register, including issues relating to the preservation of privacy can be found here.

> Furthermore, one of the major recommendations of the 2020 World Population and Housing Census Programme, launched in 2015, refers to the advantages of geo-referencing individual housing units, resulting in a housing unit register that can be later reconciled with address register.

Albeit in some developing countries, the completeness of the address register does not usually extend beyond major urban areas. The housing unit register would then be used as a major sampling frame for subsequent surveys and other statistical data collection exercises.

### 12.6.2 Conceptual framework

Even though neither a coherent conceptual framework nor an international standard for address registers exists, related concepts can be found in various places. *Handbook on geospatial infrastructure in support of Census activities* (🔗) provides guidance on organizing address information for census purposes, while the *Integrated Geospatial Information Framework* (IGIF) and its much more detailed European equivalent *INSPIRE (Infrastructure for Spatial Information in Europe) Directive* (🔗) and various national versions provide broader frameworks.

### 12.6.3 Sources and maintenance procedures

Countries that have developed administrative dwellings registers define their setup and maintenance procedures through legislative measures. A setup procedure is usually based on updated cadastral information through a specified administrative process or field recording of dwellings. The update procedure is also defined by law and includes adding and deactivating identifiers with addition/removal of every dwelling from the registry. Countries that do not have administrative dwellings and/or population registers usually construct household address registers before the census by supplementing the administrative and postal list of addresses with contact information from other sources, but there are examples where this process is performed continuously.

> Usual sources for updating the information are utility and phone companies' information, tax forms, emergency services, tv licences and similar structured databases.

The process usually involves multiple phases, where data from multiple sources is connected, cleaned and geocoded by using algorithmic tools and then controlled and corrected on the field, either by local authorities, statistical offices or external contractors. A modern approach to the process is creating a web-based address interface that enables address control and correction.

The quality of the household address register depends not only on the quality of the administrative address information but also on the quality of supplementing information. Quality of data depends on the registration process's quality but is often influenced by legal traditions, compliance and trust in institutions.

### 12.6.4 Generation of household survey frames

Household address registers can be used as directly survey frames or used to generate frames for specific surveys. A household address register is basically a list frame that lists all units in the survey population. It also contains information from an area frame, meaning information on the hierarchy of geographical areas as units. This feature of a household address register allows for two-stage sample selection, in which units are first selected from the area frame and then from the list frame only for previously selected areas.

Household address registers can be used as frames for population censuses because they contain identifying information for all census units. Further, household address register can be an excellent aid in census preparation, as they allow efficient fieldwork planning.

Unfortunately, ordinary household address registers generally do not contain information about their occupants and can therefore rarely be used for sample stratification. Population registers, on the other hand, can also serve this purpose. They can support:

- Specification of survey population;
- Derivation of survey frames;
- Selection of survey samples.
- *Fritz Scheuren - Administrative records and census taking, Survey methodology vol 25, DEC 1999* (🔗);
- *The creation of a Residential Address Register for Coverage Improvement in the 1991 Canadian Census* (🔗);
- *ONS - compiling the address register for 2011 Census* (🔗);
- *ONS – local authority partnership programme guide* (🔗);
- *Building an address index for census and beyond - Alistair Calder, ONS* (🔗);
- *Building the address register for the 2011 Census – England and Wales Guidelines on the use of registers and administrative data for population and housing censuses – UNECE* (🔗);
- *Handbook on geospatial infrastructure in support of census activities – UNSD* (🔗).

## 12.7 Household master sample, civil registration and vital statistics and population registers

Having a reliable survey frame is essential to any survey. Even though NSOs benefit from using census information to construct household survey frames, updating the survey frame is often costly and complex. Depending on available resources and administrative environments, three general approaches to household survey frames construction can be identified:

- **Use of household master sample –** commonly used by NSOs that has a sufficient number of independent surveys or periodic rounds of the same survey to sustain their use;
- **Use of population census –** used by NSOs that have sampling experts but do not have the resources or administrative sources for regular and complete updating of the survey frame;
- **Use of population registers –** used by NSOs that regularly update the complete survey frame via defined procedures from administrative sources.

Apart from the approaches mentioned above, this section will discuss the civil registration and vital statistics system, both as a source of official statistics and its usage to update the sample frame.

### 12.7.1 Household master sample

A **master sample frame** is used for sample selection in a household survey, covers the entire target population and is used for multiple surveys or multiple rounds of the same survey. NSOs usually create a master sample frame after they finish the census processing by eliminating the information that is not needed for sampling purposes and making it available to sampling experts. Depending on security and data access policies, different approaches to the technical flow of individual and contact information are followed, but the general recommendation is that access to information should be closely guarded and limited.

The main advantage of a master sample design is that it allows one or more stages to be combined or shared among different household surveys. This approach reduces the costs of developing and maintaining sampling frames as more household surveys share the same master sample frame and frame design. It also simplifies the technical process of drawing individual samples and facilitates operational linkages between different surveys. Being a master sample frame, the sampling frame does not vary either from one survey to the other or from one round to another of the same survey. Instead – and this is its distinctive characteristic – the master sample frame is designed and constructed to be a stable, established basis for selecting the sub-samples needed for the particular surveys or rounds of the same survey over an extended period of time.

Many countries face budgetary constraints and rely on technical assistance from international development agencies. To promote the sustainability of statistical data collection activities, different strategies have been proposed to economize on the technical and financial resources needed for conducting household surveys. One of these strategies is the development of a **master sample design**.

> Using a master sample design to select samples for multiple household surveys provides an opportunity to minimize the resources needed to collect household survey data regularly.

In particular, using the same sampling design and frame to select samples either for multiple surveys of different content or for different rounds of the same survey can produce significant cost-savings compared to developing an independent sampling design each time a household survey is carried out. Even though this strategy does not reduce the overall burden for respondents, respondents who live in a geographical area selected during the first stage may feel disproportionate response burden, particularly as surveys will not be collected in other areas.

> A master sample is a sample drawn from a frame covering a target population for use on a number of future occasions, so as to avoid ad hoc sampling on each occasion. Sometimes the master sample is large, and subsequent inquiries are based on a sub-sample taken from it. The main advantage of using a master sample

is that it lowers the cost of sampling as it provides a well-defined set of statistical units, which can be suitable, for instance, for panel surveys.

On the other hand, a master sample is a preselected set of units, which might cause bias when used as the source for a sample survey. Countries with good registers are increasingly moving to samples of individuals and use telephone or web interviewing in combination. The benefits of a master sample are not obvious in that case if the sources are regularly updated, and the sampling expertise is available in-house.

The terms 'master sample frame' and 'master sample' can be confused, especially if a multi-stage sampling design is involved. The defining characteristic of a master sample frame is that it serves as a sampling frame for several samples; for different surveys or different survey rounds. The defining characteristic of a master sample is that it is a sample, and different sub-samples (for different surveys or different survey rounds) are drawn from it. It is important to keep in mind that not every sample drawn from a master sample frame is a master sample.

In general, a master sample consists of an initial selection of primary sampling units (PSUs) that remain fixed for each subsample. Note that the latter stages are usually variable. For example, in the final stage of selection, the particular households chosen for interviews are usually different for independent surveys, while they may be the same or partially overlapping in repetitive surveys.

- *National household survey capability programme - Sampling Frames and Sample Designs for Integrated Household Survey Programmes* (🔗);
- *Hans Pettersson - Design of master sampling frames and master samples for household surveys in developing countries* (🔗).

***Sources, creation and maintenance procedures***

As noted in *Chapter 9.2.4 — Survey design*, in *Chapter 12.3 — Statistical Business Register* and in *Chapter 12.4 — Frames for informal sector surveys*, a business survey frame is derived from an SBR or in the absence of that, an economic census and the same applies to master sample frames for household surveys. A consistent survey frame is recommended for surveys with the same target population or a subset of the target population.

From a master sample frame, it is possible to select the samples for different surveys entirely independently. However, in many cases, substantial benefits result from selecting one large sample, a master sample, and then selecting sub-samples of this master sample to service different (but related) surveys. Several NSOs have decided to develop a master sample to serve the needs of their household surveys.

Over the time period during which a master sample frame is in use, changes will occur that affect both the number and the definition of the frame units. A master sampling frame is basically a list of area units that cover the whole country. For each unit there may be information on urban/rural classification, identification of higher-level units (administrative subdivisions – i.e., district and province to which the unit belongs), population counts and other characteristics. For each area unit, there must also be information on the boundaries of the unit. The amount of work required to maintain a master sampling frame depends primarily on the stability of these frame units. There are two kinds of changes that may occur in the frame units: changes in frame unit boundaries and changes in frame unit characteristics. Changes affecting the boundaries of frame units must be recorded in the master sample frame, and a system for collecting information about administrative changes needs to be established to keep track of these changes. Changes in frame unit characteristics include simple changes such as name changes and more substantial changes like changes in the measure of size (population or number of households/dwelling units) or changes in urban/rural classification. These changes do not necessarily have to be reflected in the master sample frame. However, as has been said above, outdated information on measures of size results in a loss of efficiency in the samples selected from the frame. Updating measures of size for the whole frame would be very costly and generally not cost-efficient; but for especially fast-growing peri-urban areas, it is a good idea to update the measures of size regularly.

There are examples where a master sample frame is periodically or continuously partially updated from CRVS systems (if the master sample contains individual information) and other administrative sources (such as local government registers, lists of voters, lists of immigrants and emigrants…) to reflect deaths, births and migrations. Also, there are examples where the update of the master sample is performed through updating address information from administrative information or via area canvassing. This is rather expensive and therefore rarely used. Examples where canvassing are focused on newly

constructed buildings identified from geospatial (or other) information. Even though that periodic updates are beneficial, a full update of the master sample frame should be performed as soon as census microdata becomes available.

***Generation of survey frames and samples***

As noted earlier, a master sample is a large sample from which different sub-samples are drawn for different surveys. The subsampling can be carried out in many different ways. Subsampling on the primary level (of primary sampling units) would give a unique subsample of the master sample primary sampling units for each survey, as each survey would have a different sample of enumeration areas. Subsampling on the secondary level would give a subsample of housing units from each master sample primary sampling units, as each survey would have the same sample of enumeration areas but different samples of housing units within the enumeration areas.

The subsampling could be carried out independently, or some kind of controlled selection process could be employed to ensure that the overlap between samples is at a desired level. Another way of selecting samples from the master sample is to select independent replicates from the sample. One or several of the replicates could be selected as a subsample for each survey. Such a set-up would require that the master sample be built up from the start from a set of fully independent replicates.

- *ST/ESA/STAT/SER.F/98 - Designing Household Survey Samples: Practical Guidelines* (🔗);
- *Maligalig, D. S., & Martinez, A. Jr (2013). Developing a Master Sample Design for Households Surveys in Developing Countries: A Case Study in Bangladesh* (🔗);
- *ESA/STAT/AC.93/3 - Sampling frames and master samples - Anthony G. Turner* (🔗);
- *UNSD - Household Sample Surveys in Developing and Transition Countries* (🔗);
- *ESSnet KOMUSO – Quality Guidelines for Frames in Social Statistics* (🔗).

### 12.7.2 Civil registration and vital statistics systems

Ideally, Civil Registration and Vital Statistics (CRVS) is the continuous, permanent, compulsory and universal recording of the occurrence and characteristics of vital events of the population. The actors in a CRVS system typically include the civil registration authorities, Ministry of Interior or Home Affairs, Ministry of Justice, Ministry of Health and the national statistical office.

> CRVS provides an important administrative function as it records documentary evidence of life events and family relations. Functioning CRVS systems record every birth, adoption, marriage, divorce and death, alongside additional information – such as the cause of death.

Recording every vital event is also an opportunity to produce the most accurate, complete and timely statistics on the population's health and demographics and maintain a national civil registry or a population database. While producing vital statistics through sample surveys serves a valuable function of highlighting and drawing attention to the possible scale of the situation, these can be costly, uncertain and may not cover the most vulnerable segments of the population, nor can they be localized. They are not a long-term substitute for vital statistics derived from a universal CRVS system. Even though CRVS systems are in place in most of the world, a substantial part of the world population did not get registered at birth in a civil registration system. UN Statistics Division has initiated the formation of the Global Civil Registration and Vital Statistics Group (🔗), a group of international and regional organizations to strengthen and support the establishment and modernisation of national CRVS systems. Similarly, Statistical division of United Nations ESCAP has launched a CRVS promotion campaign Get into the picture (🔗) that is promoting the establishment of CRVS systems in Asia and the Pacific.

> At the end of May of 2019, the United Nations Legal Identity Agenda was launched in Rome on the Conference in preparation for High-Level Political Forum: Peaceful, Just and Inclusive Societies – SDG 16 implementation and the path towards leaving no one behind. The United Nations Legal Identity Agenda refers to the holistic approach to civil registration of all vital events, production of vital statistics, the establishment and maintenance of population registers and identity management apparatus from birth to death,

> and there should be full interoperability[1] between these functions in a simultaneous manner, according to international standards and recommendations and in compliance with human rights of all people concerned, including the right to privacy. All Member States should adopt and implement this agenda as a systematic and perpetual mechanism for ensuring legal identity for all. The United Nations Statistical Commission adopted the UN LIA at its 51st Session in March 2020.

Apart from their use as a source of statistics on demographic events, CRVS systems (in combination with census and other sources) can be used to produce population estimates in inter-census periods and also population projections. Further, individual information from CRVS systems can be used to update the master sample frame information. Updating the master sample frame (as well as master sample) from CRVS systems is recommended, particularly for removing the deceased persons from the survey frame as this significantly reduces inconveniences for interviewers and consequently for NSOs.

- *Principles and Recommendations for a Vital Statistics System, United Nations, 2015* (🔗);
- *Handbook on Civil Registration and Vital Statistics Systems: Management, Operation and Maintenance Revision 1, United Nations, 2018* (🔗);
- *Guidelines on the Legislative Framework for Civil Registration, Vital Statistics and Identity Management, United Nations, 2019 Population registers* (🔗).

Having a fully reliable CRVS system and using its individual information to update the sample frame is not enough to get the complete sample frame in line with the total population. Even though information on life events is present, CRVS generally does not track information about the residence of individuals. Even if such information is present in CRVS or in a different administrative database, it is often not regularly updated or does not contain links between a person and a dwelling, thus making it difficult to update the individual information on households from the sample frame.

In some (mainly northern European countries, also Slovenia and Austria) individuals are obliged by law to register themselves administratively to a specific dwelling, and that information along with items (such as educational attainment, and current economic activity) is stored in an administrative population register. An administrative population register is usually, connected with dwellings and buildings register (described in *Chapter 12.6 — Household address register*) on individual level, as each unique person (with unique personal ID) has to be connected with a unique dwelling (with unique dwelling ID).

> Having a reliable administrative population register greatly reduces the total costs of official statistics as it eliminates the need for field collection of information that is available from reliable administrative sources.

Countries with administrative population registers do not perform field collection of Census data but connect different administrative databases and may conduct a survey to collect information that enables imputation of missing data. Even though this approach is cost-effective and enables the annual production of otherwise available indicators every five or ten years, administrative data holders are generally concerned about the response burden, which generally leads to a limited number of available variables being collected. Further, the transition to administrative census opens the problem of comparability between previous censuses.

In the context of the United Nations Legal Identity Agenda (described above), the role of population registers extends well beyond both governmental and statistical functions and becomes a mechanism for ensuring the legal identity for all from birth to death by issuing legal credentials based on civil registration throughout the lifetime of an individual. The population register – accurate and updated regularly – represents a reliable source of information on population migration. In the context of generating small area census statistics, the establishment, maintenance and operations of a universal population register – combined with information synthesized from other functional registers – represents a significant and efficient alternative to costly canvassing of the whole country.

Consequently, it is of particular importance for NSO's to be closely involved in the legislative process, establishment, operations and protocols to access the information from population registers for statistical purposes from the onset of such a process in countries. While establishing the statistical exploitation of population registers at the later stage would

[1] Interoperability in this context refers primarily to ensuring that systems are using the same set of definitions, classifications and methodology, as well as technologically compatible platforms allowing for fully harmonization of interfaces and access protocols. Interoperability between functions does not infer allowing full access and manipulating records and content of any single system.

not be impossible, it certainly would require more adjustments than if those relationships are elaborated at the very beginning.

## 12.8 Methodological services

> This section contains information about the methodological services that are part of the common statistical infrastructure and that support questionnaire design, sample design and estimation, editing, coding, imputation, outlier determination, seasonal adjustment, time series analysis, and confidentiality and statistical disclosure control activities associated with a survey or other statistical process.

Methodological services involve the use of **specialised tools and systems**, and of **specialised staff (typically methodologists)** who are expert in the design, development or acquisition, and use of such tools and systems. In this context, a tool implies a computer application supporting statistical activity, and a system is an integrated set of tools supporting a range of statistical activities.

International standards provide the foundation for developing methodological services and international statistical organizations provide support in their application.

### 12.8.1 Questionnaire design

***Provision of questionnaire design support***

Responsibility for the design and development of a survey questionnaire lies with the subject matter area staff for the survey. Questionnaire design specialists should provide support for questionnaire design and development, typically located in a methodology unit. Support may include:

- identification and acquisition of one or more questionnaire design and development tools;
- training of NSO staff in the questionnaire design principles, practices and use of the tools;
- assistance to subject matter staff in the design and testing of questionnaires;
- review of questionnaires from the perspective of understandability and question flow, and suggestions for improvements.

The principles on which the services are based are described in *Chapter 9.2.5 — Data collection and capture modes*.

***Questionnaire design guidelines, tools and systems***

Many NSOs have developed questionnaire design guidelines to assist their staff. For example Questionnaire Design (🔗) is a section within the *Basic Survey Design Manual* (🔗) developed by the Australian Bureau of Statistics.

In addition to guidelines, there are many questionnaire design tools available from NSOs, international statistical organizations, and commercially. Most of these tools are part of larger systems that, in addition to questionnaire design, include data collection and capture, editing, imputation and tabulation.

Two of the best-known systems from NSOs are:

- Blaise developed by Statistics Netherlands, and which supports questionnaire design and all types of computer-assisted data collection; and
- Census and Survey Processing System (CSPro) developed by the US Census Bureau and which is a public domain software package used by hundreds of organizations for entering, editing, tabulating, and disseminating census and survey data.

Two of the best-known systems from international statistical organizations are. Survey Solutions (🔗), free software developed in the Data Group of the World Bank, and EUSurvey, an online survey management system for creating questionnaires.

Systems from the commercial world include:

- Survey123 for ArcGIS, which (🔗) is a form-centric data gathering application that integrates the use of mapping technology and survey operations;
- Google forms (🔗), which is free and supports an unlimited number of surveys each with an unlimited number of respondents;
- SurveyMonkey (🔗) is similar to Google Forms in that it supports any kind of online survey. The free version supports only a very small number of respondents.

All these systems are further discussed in *Chapter 15.8 — Questionnaire design tools*.

### 12.8.2 Sample design and estimation

***Provision of sample design and estimation support***

The close relationship between the estimation scheme and the sample design is often designed together even though estimation actually takes place much later in the survey process than sample selection.

> It is broadly acknowledged that sample design and estimation should be fully delegated to specialists in these subprocesses as they require more mathematical knowledge than other survey subprocesses. Typically, but not invariably, the specialists are located in a methodology unit. Sometimes they are embedded within the subject matter areas responsible for the surveys.

Responsibility for the design of, development/acquisition of, and support in use of sample design and estimation tools virtually always rests with specialists located in a methodology unit. For any given survey, the subject matter staff are responsible for specifying requirements and constraints in terms of sample size, planned output tables, acceptable sampling errors, data collection budget and costs, etc., and for checking that the resulting sampling and estimation methods satisfy these requirements and constraints. The support services typically include:

- identification of appropriate sampling and estimation procedures;
- identification and (if need be) acquisition of appropriate sampling and estimation tools;
- conduct of sampling terminating with verification of the final sample with subject matter area;
- support to subject matter area in conduct of estimation and interpretation of sampling errors.

The principles on which the services are based are described in *Chapter 9.2.4 — Survey design* and *Chapter 9.2.6 — Processing survey*.

***Sample design and estimation guidelines***

Many NSOs have developed sample design and estimation guidelines for their staff, both subject matter and methodology experts. An example is Sample Design, Australian Bureau of Statistics (ABS), which is a section within the *ABS's Basic Survey Design* (🔗) documentation. It deals with:

- defining the population, frame and units;
- calculating the sample size;
- determining the sampling methodology; and
- determining the estimation method to be used.

The document includes a review of non-probability sampling methods, including quota sampling, convenience and haphazard sampling, and judgement (purposive) sampling, and the circumstances in which they might be used. It describes

simple random sampling with and without replacement, systematic sampling, stratified sampling, sample allocation, cluster sampling and multi-stage sampling, post-stratification and the circumstances within which each of these might be appropriate.

In addition to guidelines developed by NSOs, there are many textbooks and research articles on design and estimation. A classic textbook is *Sampling Techniques, Cochran, Third Edition, 1977, Wiley* (🔗); and a classic article is *Sampling and Estimation for Establishment Surveys, 1994, M A Hidiroglou* (🔗).

The journals published by the International Statistical Institute (ISI), specifically including the Journal of Official Statistics (🔗), and the various national association such as the American Statistical Association (ASA), and the Royal Statistical Society (RSS) are rich sources of articles.

***Sample design and estimation tools and systems***

In the past, an NSO would develop tools for stratification, sample size determination, sample selection, and estimation itself, often separately for each survey. This is no longer common practice as tools for every aspect of sample design, and estimation are readily available commercially and from international organizations and other NSOs. Typically, these tools are combined with one another and with tools for data preparation, analysis and tabulation in a single system.

Three of the best-known commercially systems are listed below. Purchase of any of these systems must be accompanied by training. All have a broad range of features, some of which may be complex and not all of which may be relevant to an NSO.

- SAS is a software suite that can discover, alter, manage and retrieve data from various sources and perform statistical analysis on them. It provides a graphical point-and-click user interface for non-technical users and more advanced options through the SAS language.
- SPSS Statistics is a statistical software platform from IBM with essentially the same features as SAS.
- Stata is a statistical software platform with essentially the same features as SAS and SPSS.

There are many freely available systems covering a wide range of functions of which. R (Project for Statistical Computing) is the best known. Although the system is free, the NSO staff require training in its application, which has to be purchased.

Over one hundred systems are available through GitHub.

The systems referenced above are described in *Chapter 15.7 — Specialist statistical processing/analytical software*.

### 12.8.3 Editing, coding, imputation and outlier determination

***Provision of support for editing, coding, imputation and outlier determination***

The ultimate responsibility for executing these activities for any particular statistical production process typically lies with the subject matter area manager. However, in most NSOs, primary editing and (sometimes) coding are actually carried out by the staff who undertake data collection, typically field staff working from regional offices. Secondary editing, imputation and outlier determination may be carried out by specialists in these activities located in a methodology unit or in the relevant subject matter area.

> Responsibility for the design of, development/acquisition of, and support in the use of, generic editing, coding, imputation and outlier detection tools and systems typically rests with specialists located in a methodology and/or ICT units.

In summary, support for editing, coding, imputation and outlier determination may include:

- identification and/or development of appropriate procedures and tools;
- training of staff in the editing, coding, imputation and outlier detection procedures and tools;
- assistance to staff in the conduct of these activities.

The principles on which the services are based are described in *Chapter 9.2.6 — Processing survey*.

***Editing, coding, imputation and outlier determination guidelines, tools and systems***

Many NSOs have developed questionnaire design guidelines for these activities to assist their staff. For example, Data Processing (🔗) is a section within the *Basic Survey Design Manual* (🔗) developed by the Australian Bureau of Statistics. Generic Statistical Data Editing Models (🔗) have been developed by a multinational task team under the High-Level Group for the Modernisation of Official Statistics (🔗).

In many NSOs the editing, coding, imputation and outlier determination tools are built separately for each production process (survey or administrative data collection) and, in the case of annual or less frequent surveys/collections for each cycle. This is not recommended practice. To the extent possible, the best approach is to use generic tools that can be customised to a particular survey or collection. Tools may be developed in house or, preferably, acquired from another NSO or an international statistical organization. In some cases, tools performing more than one of the functions may be combined in a system.

Systems currently available include CSPro, Survey Solutions, and Blaise, as noted above and described in *Chapter 15.7 — Specialist statistical processing/analytical software*.

## 12.8.4 Seasonal adjustment and time series analysis

***Provision of support for seasonal adjustment***

As discussed in *Chapter 10.3.1 — Methods of analysis*, seasonal adjustment is a method widely used in official statistics for removing the seasonal component of a sub-annual (usually monthly or quarterly) time series. It includes pre-treatment, which involves detection and correction of outliers and calendar adjustment, i.e., removing trading day variations and moving holiday effects. In some cases, the original series may be differenced, i.e., a new series derived that comprises the differences between adjacent points in the original time series. The various choices made in setting up a seasonal adjustment (including pre-treatment) for a particular series are collectively referred to as model selection.

Responsibility for the seasonal adjustment rests with the subject matter area responsible for the series. The model selection and seasonal adjustment algorithms are complex and depend upon knowledge and use of a seasonal adjustment system. Therefore, the subject matter staff is typically helped by seasonal adjustment specialists, who are usually located within a methodology unit or an analysis unit, depending upon the organizational structure of the NSO. Their role is:

- to decide upon the seasonal adjustment system(s) to be used across the NSO as a whole;
- to acquire and install the seasonal adjustment system(s), to test it and to adjust it for each time series;
- to determine the appropriate approach and system (if more than one is available);
- to establish the initial seasonal adjustment model and parameter settings for the series;
- to check that the model and settings work appropriately on the series before handing over the system to the subject matter area; and
- periodically, to review the outputs, check the continuing applicability of the model and settings, and to make adjustments if need be.

Support may also be provided from outside the NSO. For example, Eurostat provides a free seasonal adjustment remote helpdesk (🔗).

***Seasonal adjustment guidelines, tools and systems***

*ESS Guidelines on seasonal adjustment, 2015, Eurostat* (🔗) provide a comprehensive description of all aspects of seasonal adjustment, including pre-treatment and model selection.

There are multiple seasonal adjustment systems available, of which the most commonly used are listed below and described in *Chapter 15.7 — Specialist statistical processing/analytical software*.

- X-12 ARIMA, US Census Bureau.

- TRAMO-SEATS, Department of Statistics, National Bank of Spain.
- X-13 ARIMA-SEATS system, which combines X-12 ARIMA and TRAMO-SEATS, developed and supported by the US Bureau of the Census.
- Jdemetra+, also combines X-12 ARIMA and TRAMO-SEATS, developed by the Department of Statistics, in the National Bank of Belgium for the ESS Seasonal Adjustment Group.

It is highly recommended that an NSO use the same seasonal adjustment system for all the series that are seasonally adjusted. This allows staff to become familiar with the system. In any case, a single system may allow more than one approach to seasonal adjustment. To the extent possible, the same approach should be used for all series. However, there may be a case for using different approaches in different domains.

### 12.8.5 Confidentiality and disclosure control

***Provision of support for confidentiality and disclosure control***

As discussed in *Chapter 3.2.6 — Principle 6 - Confidentiality*, Principle 6 of the Fundamental Principles of Official Statistics (🔗) states:

> "Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes."

Confidentiality is ensured through measures such as:

- protecting questionnaires during data collection and when in transit – as discussed in *Chapter 15.2.17 - Data security*;
- requiring all employees to swear an oath not to disclose confidential information;
- restricting access to buildings and servers with confidential information – as discussed in *Chapter 16.3 - Building security*;
- implementing confidentiality checking and disclosure control procedures – as discussed in *Chapter 10.3.1 - Methods of analysis*.

The first three items above are security issues. This section focuses on the fourth item, support for confidentiality checking and disclosure control.

Ensuring that there is no disclosure of confidential data in the output tables is the responsibility of the subject matter staff responsible for the statistical production process. However, these staff may not have the specialized skills required for confidentiality checking and preservation and may well draw on specialist support in selecting and using the appropriate tools. The specialists are usually located in methodology unit or an analysis unit, depending upon the NSO's organizational structure. The support services typically include:

- specification of appropriate confidentiality checking and disclosure control procedures and identification and (if need be) acquisition of a corresponding confidentiality checking and disclosure control tool, for use throughout the NSO;
- and, for each set of output tables from a statistical production process:
- support to subject matter area in their conduct of confidentiality checking and disclosure control; and
- periodic verification of confidentiality preservation effectiveness.

***Confidentiality checking and disclosure control tools***

Identifying and preventing disclosure is not a process that can readily be done, or should be done, manually. First, the tables are typically too complicated and/or voluminous. Second, manual processing is inefficient when the job can be much more readily done by automated processing. Thus, an NSO should either acquire a confidentiality checking and prevention tool as an element of its common statistical infrastructure or develop its own tool. Acquisition is recommended wherever possible to save development costs and to be more certain that the tool does the job properly. However, as confidentiality checking, and preservation tools are not readily available commercially (there being very little demand for

such tools outside realms of official statistics) acquisition is likely to be from another NSO. Two well-known examples are as follows:

- **ARGUS, Statistics Netherlands**

  As described in ARGUS Users' Manual Version 3.3 (🔗), the purpose of Γ-ARGUS is to protect tables against the risk of disclosure. This is achieved by modifying the tables so that they contain less detailed information. Several modifications of a table are possible - a table can be redesigned, meaning that rows and columns can be combined and/or sensitive cells can be suppressed and together with any additional cells required to protect against residual disclosure. Cells to protect these can be found in some optimum way (secondary cell suppression). A twin application, μ-ARGUS protects microdata files. Both applications have been rewritten in open source (🔗).

- **G-Confid, Statistics Canada**

  As described in G-Confid: Turning the tables on disclosure risk, 2013 (🔗), cell suppression is the technique used to protect tabular economic data in the disclosure control application G-Confid, developed at Statistics Canada. It is a generalized system that can deal with potentially voluminous multi-dimensional tables and incorporate new approaches. Its main objective is to provide the appropriate protection level for confidential cells while minimizing the loss of information resulting from the process.

  G-Confid features a suite of three SAS components for use with tabular economic data at various aggregation levels. PROC SENSITIVITY identifies cells requiring primary suppression. The macro SUPPRESS protects the cells identified by PROC SENSITIVITY by selecting an optimal set of cells for complementary suppression using a linear programming algorithm. The macro AUDIT validates a suppression pattern not provided by the macro SUPPRESS or the G-Confid user altered after running the macro SUPPRESS. An additional macro, AGGREGATE, provides further information about sensitive unions of cells, and another auxiliary macro REPORTCELLS provides a visual snapshot of the suppression pattern to facilitate the creation of output tables of the economic data under study.

## 12.9 Statistical policies, standards and guidelines

The final aspect of the statistical infrastructure discussed in this chapter comprises statistical policies, standards and guidelines, defined as follows:

- **Policy** –strategic direction or course of action defined, adopted and promulgated by the NSO;
- **Standard**: set of standard concepts or variables, standard classifications, or standard method that underpins harmonisation and/or integration and that typically supports the implementation of policies;
- **Guideline**: a recommendation and/or element of best practice that typically supports the implementation of a policy or standard; usually used in the plural, referring to a set of guidelines on a particular topic.

### 12.9.1 Statistical policies

Whilst a **principle** is a statement of commitment (possibly a rule) that is to be followed, a **policy** defines a course of action that can be used to implement the principle. Thus, an NSO may adopt the Fundamental Principles of Official Statistics, and put in place one or more policies, such as a dissemination policy, to ensure that they are followed. This section focuses on national statistical policies, which are most commonly used to elaborate and provide strategic directions for implementing the principles or rules defined in legal documents and international recommendations.

> Policies should be like the foundation of a building - built to last, and resistant to change and erosion. Policies should be easily accessible and understood. Policies are usually developed to provide a sense of direction in a strategically important area. They should be created to be in force for a long period and at the same time periodically reviewed to assess their relevance.

Policies may be divided into three types: **organization-wide; issue-specific; and system-specific**.

Commonly NSOs have policies for dissemination, revision, data access, microdata access, archiving, privacy, pricing for tailored statistical services among many others.

Recommendations on how to organize the development, monitoring and enforcement of policies is provided in *Chapter 6.4.2 — Internal advisory and decision-making bodies*. One general recommendation is that policies should be created through a designated internal decision-making body (such as a committee or a working group), reviewed by senior experts and management, and endorsed by the chief statistician.

Senior management should give special attention to policy enforcement, continually reminding the staff of the importance of a policy and why it was adopted. Policy performance monitoring should be continuous activity as ad hoc performance monitoring can create unreliable results.

Enforcement of statistical policies may be organized globally, for example through the *Global Review of the Implementation of the Fundamental Principles of Official statistics* (🔗), or regionally for example through the Peer Review on the implementation of the European Statistics Code of Practice (🔗) in the EU and candidate countries.

Statistical policy governance should involve eight characteristics: the rule of law; transparency; responsiveness; consensus-oriented; equity and inclusiveness; effectiveness and efficiency; accountability; and participation.

An example of good practice for statistical policy creation can be found in the UK Government Statistical Service (GSS) document.

*Governance for statistical policies, standards and guidance statistical system* (🔗), which summarises the process for creating, updating and agreeing on all GSS. Policies, standards, guidance and strategies. All GSS policies are made available on the GSS Policy and Guidance Hub, which has the following objectives:

- helping people to find what they are looking for;
- increasing awareness of existing policy;
- preventing use of out-of-date policy;
- avoiding unnecessary new policies / reinventing the wheel; and
- identifying gaps in policy areas.

In the Philippines Statistical Authority (PSA), interagency committees on statistics (IACs) are established to serve as fora for the exchange of views and expertise:

- to resolve technical issues and problems arising from the production, dissemination, and use of statistics, among others and;
- in the development, maintenance and compliance to statistical standards and classification systems.

Statistical issues that are unresolved at the IAC level are elevated to the PSA Board, the country's highest policymaking body on statistical matters, for decision. Policies issued by the PSA Board are then implemented by all concerned agencies.

- *CSO Ireland - Standards and Guidelines* (🔗);
- *US Office of Management and Budget - Standards and Guidelines for Statistical Surveys* (🔗).

### 12.9.2 Standard concepts, variables and classifications

Today's world requires comparable statistics. Political unions, customs unions, free trade areas, and the global character of foreign investment require that statistics (and particularly socio-economic statistics) are expressed in the same language. The objective of harmonising statistics is to make it possible to combine or compare data collected for different populations, for different periods and/or by different data collection methods for different statistical units. This is possible by using the same or mutually consistent and harmonized standards and classifications across different data sets.

A failure to use internationally comparable standards is a weakness in an NSO. For effective coordination of classifications, nomenclatures and concepts, a designated organizational unit must be in charge of ensuring that what is adopted nationally is comparable with that adopted by other countries.

> International comparisons are possible largely due to international classifications and standard accounting systems such as the System of National Accounts (SNA) or System of Environmental Accounts (SEA). In efforts to increase comparability, the statistical community is also developing standards such as GSBPM, GSIM, and CSPA (described in *Chapter 15.4 — Use of standards and generic models in an NSO*) aimed at standardising the statistical production and its documentation.

Classifications group and organize information meaningfully and systematically into a standard format that is useful for determining the similarity of ideas, events, objects, or persons. The preparation of a classification means creating an exhaustive and structured set of mutually exclusive and well-described categories, often presented as a hierarchy reflected by the numeric or alphabetical codes assigned to them.

International statistical classifications (ISCs) are products of international agreements among national authorities responsible for statistics in the respective areas. In accordance with the established practice for the division of responsibilities among international agencies in the area of statistics, the ISCs require approval by the United Nations Statistical Commission (UNSC) or another competent intergovernmental board, such as that of the World Customs Organization (WCO), the World Health Organization (WHO), the International Monetary Fund (IMF), or the International Labour Organization (ILO), depending on the subject matter area. ISCs may serve as models for the development of corresponding national, multinational and regional statistical classifications, and should, as far as possible, reflect what is considered "best practice" in the areas they cover.

Even though international standardisation of variables was not until recently seen as a priority, standardisation of data and metadata has been promoted through international organizations such as the IMF and the World Bank, the OECD and the EU. All of them accept data in their databases through XML and require standardised variables and metadata. This simplifies the use of data and makes data integration easier.

Within the European Statistical System (ESS), an **ESS standard** is defined as a normative document, established by consensus among ESS members and approved by a recognised body for ESS standardisation, that provides common and repeated use by several actors in the ESS. It can comprise rules, guidelines or characteristics for the development, production and dissemination of European Statistics, aimed at achieving the optimum implementation of the mission and vision of the ESS.

Adapting the statistical classifications to local circumstances and conditions and promoting their use among public and government agencies is an important task for every NSO.

- Catalogue of ESS Standards (🔗);
- *Standard Statistical Classifications: Basic Principles* (🔗);
- Definitions, data sources and methods – Statistics Canada (🔗);
- *Best Practice Guidelines for Developing International Statistical Classifications -, Expert Group on International Statistical Classifications* (🔗).

### 12.9.3 Statistical guidelines

> In the world of official statistics, guidelines are used to document the best current practices and processes and explore and promote the new developments.

Even though guidelines are, by nature, open to interpretation and do not need to be followed to the letter, they are worth developing, promoting and considering.

Guidelines cover every statistical area and domain. They provide a basis for training and support. When a new employee is introduced to an NSO, he/she can be given a copy of pertinent guidelines as a starting point. He/she may be asked to follow the work of more senior colleagues and write down observations and questions with reference to the guidelines. Since the guidelines are usually written following the statistical production process from the start to the end, they can be used as supporting tools to troubleshoot errors in streamlining and improving existing processes.

Creating guidelines in an NSO is somewhat similar to creating an internal policy as referenced above. Guidelines should be created through a working group or committee of experts, then commented on by internal decision-making body

(such as a Committee or a working group), then reviewed by senior experts and management and finally endorsed by the chief statistician. In the case of guidelines being developed by an international organization, the final approval is by a commission or another competent intergovernmental body.

As international guidelines are often not mandatory, compliance is in general, monitored and not audited. International organizations often monitor compliance with international guidelines, for example, for EU and OECD applicant countries during the accession process.

- *CSO Ireland - Standards and Guidelines* (🔗);
- *US Office of Management and Budget - Standards and Guidelines for Statistical Surveys* (🔗);
- *ISTAT - Guidelines for the quality of statistical processes that use administrative data* (🔗).