Draft Guidelines on the statistical infrastructure required to support the production of business demography and entrepreneurship statistics

Chapter 1: Introduction

1.1 Measuring entrepreneurship is important

The role of entrepreneurs in stimulating economic growth has received renewed attention in the past two decades, when the extraordinary development of information and communication technologies and other advanced technologies fostered new breeds of start-ups, as well as a new attitude toward entrepreneurial activity. The interest in entrepreneurship was further stimulated by the economic recession that followed the global financial crisis. The notion that entrepreneurship is a crucial driver of economic growth has become part of the policy narrative on how to resume growth and prosperity across the world.

Though, the theoretical foundations of such an assumption are, to some extent, still under development. While the work of Schumpeter at the beginning of the 20th century had already identified entrepreneurship as a driving force for innovation and an engine for economic development, it was only in the past two decade that some scholars proposed to formally consider entrepreneurship as an additional factor in explaining economic growth. The assumption is that entrepreneurship is distinct from human capital and R&D, two drivers of growth already put forward by endogenous growth theories, and that it constitutes the missing link between investment in new knowledge and economic growth: entrepreneurs, via firm creation, exploit opportunities provided by new knowledge and ideas that are not yet discovered and commercialised by incumbent firms.

A second mechanism is also at work. The Schumpeterian process of "creative destruction" itself is a driver of economic growth: new firms entering the market displace obsolescent firms, and the business dynamics of entry and exit that results from it contribute to the productivity dynamics and eventually to economic growth.

Against this positive backdrop, the literature on entrepreneurship also acknowledges that entrepreneurs compose a heterogeneous group that encompasses very different types. Innovative "Schumpeterian" entrepreneurs coexist with "necessity entrepreneurs", who create new businesses out of necessity and not to exploit market opportunities or innovative ideas. Not all entrepreneurial activities will then have the same, positive impact on productivity and growth; entrepreneurship will be more or less conducive to knowledge creation and diffusion and economic growth depending on how and where it occurs, e.g. in which sectors and/or location.

In fact, the impact of entrepreneurship has also been associated to poverty reduction and social inclusion. Stimulating entrepreneurial activities among individuals who are part of disadvantaged or marginalised groups (because of their age, gender, ethnic characteristics, or lack of education and skills, or the geographical area where they live) can be an important vehicle of social inclusion and contribute to poverty reduction by bringing into a country's active labour force segments of the population previously excluded.

To advance understanding of the role of entrepreneurship in economic growth and social inclusion, sound evidence on the entrepreneurial phenomenon, its determinants and impacts is critical, but today empirical research on entrepreneurship continues to rely to a large extent on private data sources and much less on official statistics.[1] Indeed, suitable and comprehensive data for the analysis of entrepreneurship not always are found in national statistical offices (NSOs), despite the wealth of economic statistics on businesses they produce.

[1] At the international level, three collections of entrepreneurship data developed outside NSOs have gained popularity among analysts: the Global Entrepreneurship Monitor (GEM), the Eurobarometer on Entrepreneurship of the European Commission, and the World Bank Entrepreneurship Databases. The first two collections rely on ad-hoc surveys of individuals, while the third mainly consists of administrative data on business registrations.

The present *Guidelines* aim at assisting NSOs and other statistical institutions in the production of statistics on entrepreneurship. The main challenge is that entrepreneurship data cross the traditional boundaries of economics and social statistics, by encompassing topics such as the demography and performance of businesses as well as the profile of individuals who create enterprises. These *Guidelines* focus on economic statistics and illustrate concepts and methods to generate statistics on the demography of businesses to complement the wealth of business data already produced by NSOs, in particular structural business statistics. The latter have existed for some time to describe the structural characteristics of the business population (e.g. business counts, employment, value added, turnover, etc.) but cannot adequately account for the entrepreneurial dynamism as reflected by the creation of new businesses, their survival, grow or death. Business demography statistics, on the contrary, capture these fundamental dimensions of the phenomenon of entrepreneurship.

1.2 Statistics for measuring entrepreneurship

The topic of entrepreneurship was included as a statistical area in the Classification of Statistical Activities (CSA) only recently, in 2010.[2] Specifically, entrepreneurship is categorised under Domain 3: Environment and multi-domain statistics. As statistical area 3.3.7, entrepreneurship statistics cover "the measurement of the determinants, performance and impact of entrepreneurial activities of people and organizations". This classification acknowledges that entrepreneurship statistics are a multi-domain statistical area, insofar as self-employment and business demography, for example, are key elements of statistical analysis needed to gauge entrepreneurial efforts.

These *Guidelines* are concerned with measures of **entrepreneurial performance** based on business data, notably the two types described below:

a) *Statistics that focus on businesses*. These consist primarily of business demography statistics measuring the birth, death, survival and growth of enterprises as well as the employment changes associated with these events. Also, by linking business demography data with other business statistics important indicators are generated to describe the performance of new or young enterprises versus that of mature enterprises in domains such as trade or innovation. Examples of linked business data are statistics on export and innovation activity of young businesses.
b) *Statistics profiling business jointly with business owners*. These measures of entrepreneurship rely on statistics linking data on individuals who are business owners with data on the businesses they own; these statistics provide the profile of the entrepreneur together with the performance of his/her business. Progress in this area has been slower reflecting the greater compilation challenges involved. Examples are statistics on business birth and death by gender of the business owner and innovation activity of young firms by educational attainment and employment history of the business owner.

The *Guidelines* do not cover a third category of measures of entrepreneurial performance that is based on statistics on individuals only (Box 1.1).

a. *Statistics that focus on individuals* (the entrepreneurs). Entrepreneurship indicators on individuals consist of statistics on self-employment from Labour Force Surveys or Population Censuses and statistics from ad-hoc surveys of individuals on attitudes toward entrepreneurship and involvement in entrepreneurial activities. Examples are statistics on self-employed, preference for self-employment over wage employment and attitudes toward the risk of business failure.

[2] The Classification of Statistical Activities is an international standard for describing and categorising official statistical work by domain.

Box 1.1 **Entrepreneurship and self-employment**

The approach to measuring entrepreneurial activity has evolved significantly in the past two decades. Until the late 1990s, self-employment was the most frequently used indicator to proxy entrepreneurship. This represented a very convenient choice: self-employed persons are defined as persons who are the sole owners, or joint owners, of the unicorporated enterprises in which they work, and they therefore appear to well represent the figure of the entrepreneur; also, information on self-employment is collected on a regular basis in labour force surveys and population censuses, in virtually all countries, making the use of self-employment data an easy solution for analysts in search of quantitative evidence on entrepreneurship.

While self-employment data do contribute to the understanding of the entrepreneurial phenomenon, there are limits to their ability to represent entrepreneurship in a comprehensive and accurate way. Firstly, not all the self-employed are in fact entrepreneurs, e.g. some might have a self-employment status for reasons of convenience, or might be self-employed out of necessity, or might be "gig workers" providing on-demand services for an online platform in the context of a contractual relation where any notion of entrepreneurial risk is absent for the "gig workers". Secondly, self-employment can at best provide information on the population of entrepreneurs, e.g. their number, their characteristics in terms of gender, age, education or work experience, but alone cannot offer any insight about the performance of the businesses being created, for instance about their ability to survive or grow.

In light of these limits, in recent years the international statistical community made considerable efforts to conceptualise and develop additional indicators for measuring entrepreneurial activity, in particular that concentrate on businesses as observation unit instead of individuals.

1.3 Business demography and international comparability

Business demographics refer to statistics on "events like births and other creations of units, deaths and other cessations of units, and their ratio to the business population. It covers follow-up of units in time dimension, thus gaining information on their survival or discontinuity. It also covers development in time dimension according to certain characteristics like size, thus gaining information on the growth of units, or a cohort of units, by type of activity" (*Eurostat-OECD Manual on Business Demography Statistics*, 2007). To be noted, in the economic literature the term "business dynamics" is used to refer to the analysis of the demography of enterprises and its impact on employment and productivity.[3]

In principle, business demographics can be compiled from a range of data sources, in particular statistical business registers, business censuses and business surveys. Research conducted in the mid-2000s on how to improve cross-country comparability of data on the creation of new enterprises identified the following advantages and challenges associated to different sources of data on business demography (Vale, 2006):

- Statistical business registers usually provide comprehensive coverage of the population of interest and are a reliable source. In fact, data from a comprehensive, frequently updated statistical business register are likely to be more reliable than those from small scale surveys or studies. The quality of the data in the source clearly has an impact on most of the other factors of comparability identified here, for example poor quality information on economic activity will have an impact on the comparability of coverage. An issue with SBR is that

[3] Business scholars use the same term to indicate modelling methods to test the effectiveness of different policies on business outcomes and for the analysis of business strategy; for example, see J. D. Sterman (2000), *Business Dynamics: Systems thinking and modelling for a complex world*.

systematic biases in the population coverage may be present due to scope and threshold restrictions.

- Census data can be as good as statistical business registers, and sometimes better, if they have less scope restrictions. However, the cost of running a census of businesses every year makes this approach unrealistic for most countries. Data from less frequent censuses can be used, but comparability of statistics across countries will be difficult when the periodicity of data differs.

- Survey data have also been used by some countries and most notably in the Eurostat project "Demography of Small and Medium-sized Enterprises (DOSME)" conducted in twelve countries of Central and Eastern Europe at the time of their transition to the market economy. This approach is useful when registers are not sufficiently developed and allows, for instance, the collection of more information on the profile of the business creators than is available from other sources, or on the informal economy. It suffers, though, from the usual constraints of survey errors and sample size limitations when detailed data breakdowns are required. Also, it is not possible to identify business deaths through surveys. While survey data did not emerge as the best source for business demography, an important lesson was learned from the DOSME project: it demonstrated that having better data on business demography can have a strong impact on the quality of statistical business registers and more generally of all economic and business statistics. The project allowed for some of the participating countries to catch up rapidly with their western European counterparts in the quality of data on small and medium-sized enterprises (UNECE, 2013).

The review of advantages and challenges points to the use of statistical business registers as a most convenient source for the purpose of compiling international comparable business demography statistics, especially if SBRs are already subject to some form of harmonisation, i.e. if basic requirements of the registers are harmonised across countries.

Today, an increasing number of countries compile business demography statistics on a regular basis, although mostly not in a harmonised way and with important differences in the scope of the data collection (Box 1.2).

An exception is the effort conducted by numerous countries in the framework of the OECD-Eurostat Entrepreneurship Indicators Programme (EIP) launched in 2007 and, for Member States of the European Union, also to ensure compliance with mandatory regulations: these countries produce business demography statistics following the methodology and definitions recommended by the *Eurostat-OECD Manual on Business Demography Statistics* (2007). As a result, Eurostat and the OECD have created databases of internationally-comparable business demographic statistics that were previously missing in the domain of official statistics.

Box 1.2 **Overview of country practices**

In 2013, the United Nations Statistics Division (UNSD) conducted an "SBR Global Assessment Questionnaire Survey" aimed at collecting information on the status of the business register in all countries. The Global Survey Questionnaire was applicable not only to National Statistical Offices (NSOs) that operate a comprehensive single business register but also to those that maintain and update one or more lists of enterprises, economic census frames, or annual enterprise surveys. The list(s) of enterprises could be compiled based on multiple surveys or could be built by combining survey data with administrative data; for the purpose of the survey, also such lists were referred to as a business register.

The analysis of the questionnaire findings highlighted a number of facts:

- In most countries, the compilation of BD from business registers started in the early 2000s or even in more recent years. France, where the production dates back to the 1970s, represents an exception.

- In some developing and transition economies the series mentioned by the respondents consisted in structural data about the number of enterprises (by sector, size and region). This revealed a misunderstanding about the meaning of business demographics.

- Business demography statistics produced by countries based on SBRs may simply consist of data on registrations and de-registrations of firms with the competent authority (as in Singapore, for example); or, on the contrary, be the result of the implementation of ad-hoc methodology and definitions of the main demographic events. This is the case in OECD countries, EU member States and other developed and emerging economies, where either a national specific methodology (for instance, Australia, Costa Rica, the Russian Federation and Tunisia) or an internationally agreed methodology (in particular based on the *Eurostat-OECD Manual on Business Demography Statistics*) is implemented.

1.4 The aim of the *Guidelines*

The present *Guidelines* are intended to assist the construction of the statistical infrastructure for the compilation of business demography statistics and related indicators.

Guidelines and Eurostat-OECD Manual on Business Demography Statistics

The *2007 Eurostat-OECD Manual on Business Demography Statistics* made a substantive contribution to data development in this statistical area by elaborating for the first time definitions of business demographic events to serve as recommended international standards, and by proposing a methodology to operationalise the definitions and compile business demography statistics.

The current *Guidelines* complement and complete the *2007 Eurostat-OECD Manual* by further developing methodological aspects with respect to several dimensions, including: the computation of business demography at the regional level; the compilation of statistics from the linking of the statistical business register with other data sources; and methodologies for producing business demography statistics from sources different from the statistical business register.

Also the *Guidelines* emphasise the effects of the compilation of business demography in supporting improvements in the quality of business statistics and economic statistics.

Guidelines* and *UNECE Guidelines on Business Registers

These *Guidelines* also complement the *UNECE Guidelines on Business Registers* (2015), notably by addressing the requirements of the statistical business register system in order to support the production of business demography statistics. They also expand the methodological explanations and provide illustrative examples of country practices of compilation of business demography statistics from business registers.

1.5 Content of the *Guidelines*

Chapter 2: Definitions and key concepts of business demography statistics

Chapter 2 presents the key definitions of the business demography concepts and the related variables based on the Eurostat-OECD Manual on Business Demography Statistics and the Guidelines on Statistical Business Registers. The chapter provides definitions and explanations on continuity rules and the identification of entrepreneurs, including the treatment of self-employed persons. It also discusses definitions of high-growth enterprises and gazelles.

Chapter 3: Requirements of SBR to support the production of demography statistics

Chapter 3 is a core chapter focussing on how to develop SBR to facilitate the production of business demography statistics. This chapter aims at building the bridge between the concepts of business demography statistics and the necessary infrastructure in the statistical business registers in order to support the production of business demography statistics. Guidelines on this do not exist elsewhere. Development and maintenance of longitudinal businesses data plays a key role in the chapter, including methods to identify the demographic events to determine whether a business has been born, survived or died. Also other forms of register requirements are dealt with, such as coverage of the register, the updating frequency and the variables needed to describe the demographic events.

Chapter 4: Guidance for linking the SBR with information from other sources

Linking business demography data with other existing data sources is an efficient way of producing statistics and increases the analytical power of the demography data set. However, SBRs cover only a limited set of variables on the businesses and usually no information on the entrepreneur as (natural) person. The chapter provides methods and guidance on linking SBR information with other data sources. This may be data on the businesses, for instance on the success of the enterprise (e.g. growth of employment, turnover, profit) and foreign trade, and social and individual data, for example information about the entrepreneur (e. g. sex, age, nationality, (former) employment status and education) that may be obtained from surveys, administrative registers or other statistical registers.

Chapter 5: Other approaches to produce business demography statistics

While the guidelines focus on the production of business demography based on the SBR, it is recognized that countries might not be able to do so in the short run and the production of their business demography statistics is based on information from surveys. Thus, the chapter provides examples on the production of business demography statistics based on data from surveys, censuses or administrative sources or based on a combination of different data sources.

Chapter 6: Business demography statistics (OECD)

This chapter outlines a minimum set of business demography indicators recommended in order to produce comparable business demography statistics. The indicators are for convenience presented by a set of tables with explanations and recommendations.

Chapter 7: Topics for further work and research (Norbert Rainer)

The chapter includes a number of topics for possible further work and research, based on the discussions of the task force. The research agenda includes topics such as: development of longitudinal databases for producing business demography statistics; linking and integrating the SBR with information from other data sources, including administrative sources; development of a statistical framework for business demography statistics; the use of different statistical units in business demography; international comparability of business demography statistics; and sub-annual (quarterly) and regional indicators.

The guidelines also include a Glossary with definitions and explanations of key terms and concepts.

Chapter 2: Definitions and key concepts of business demography statistics

This chapter aims at providing the definitions and key concepts of business demography statistics and of statistics on high-growth enterprises. The chapter is mainly based on the Eurostat/OECD Manual on Business Demography Statistics, the Eurostat Business Register Recommendations Manual and the UNECE Guidelines on Statistical Business Registers. These manuals are complementing each other. However, the Eurostat/OECD Manual focusses on business demography and will be the main basis for this chapter. On the other hand, this Manual does not consider using other statistical units than the enterprise and also not regional business demography statistics.

Chapter 2 is organised as follows: In sub-chapter 2.1, the basic concepts of business demography are introduced, followed in sub-chapter 2.2 by a detailed explanation of the key concepts, especially concerning births, survivals and deaths. The definitions of the variables and characteristics of business demography are explained in sub-chapter 2.3. The last sub-chapter 2.4 provides the key concepts of high-growth enterprises and gazelles.

2.1 Basic concepts of business demography

2.1.1 Subject of business demography statistics

As explained in chapter 1 business demography statistics provide data on the dynamics of the business population over time and core variables that describe the development of the business population. Focus of business demography statistics is on the birth of new businesses, the survival of newly created businesses and the death of businesses. Based on these three aspects various indicators can be derived:

- Birth rate: share of newly born businesses in t as percentage of all active businesses in t
- Survival rate: share of businesses born in t that are still active in t+1, t+2, t+3, etc.
- Death rate: share of closed businesses in t as percentage of all active businesses in t

These indicators are usually broken down by economic activity (according to ISIC), legal form, size classes (employment or turnover) or regions. In case of sole proprietorship businesses, the data can be enlarged by characteristics of the entrepreneur, such as age, sex, education of the entrepreneur and further characteristics relevant for the analysis of entrepreneurship. These additional characteristics might be achieved by micro data linking with social statistics databases. Further characteristics of the enterprise can be achieved by linking business demography units with other business statistics databases, such as foreign trade database, R&D statistics and innovation statistics, to name a few (see chapter 4).

Further information on business dynamics could be the derivation of indicators for the whole population of the active businesses. One example might be data on the age of the businesses. Two important variables are used to describe the size and development of the businesses: turnover and employment. It is of special political interest to provide data on the number of jobs that were created by newly born enterprises. Of course, also the opposite needs to be taken into account: loss of jobs through cessation of enterprises.

Very much related to business demography are indicators of the growth performance of businesses. It is especially of interest to identify businesses within the whole population of active businesses that are very much growing and thus create a considerable number of new jobs. Conceptually, high-growth businesses are viewed as businesses that have an increase of employment over a certain threshold and over a certain period of time. A sub-group of high-growth enterprises are “gazelles” which are high-growth enterprises that were newly born.

Business demography statistics can be based on different statistical units and different concepts of business demography might be applied. In international data collections by the EU and OECD the statistical unit enterprise is applied. More explanation on the basic concepts is provided in the following sub-chapters.

2.1.2 Employer and economic enterprise demography

There are different concepts on how the dynamics of businesses might be depicted in business demography statistics. A first one will include all businesses, a second one just those businesses that have at least one employee, and a third one only enterprises with at least two or more employees. The first form is usually just called “business demography”, the second one “employer business demography” and the third one “economic enterprise demography”.

The concepts of births, survivals and deaths are different between these three kinds. In addition, the total populations are different as well as the derived indicators. In case of the (overall) business demography, the population is defined as those businesses that are active in a certain period (having employment or turnover). In case of employer business demography, the total population is defined as consisting only of active employer businesses with at least one employee and in case of economic enterprise demography only active businesses with at least two employees. There is usually a significant difference in the number of units, as there is a considerable high number of businesses that have no employees (self-employed businesses).

In addition, different definitions of births, survivals and deaths do apply. For more details, see sub-chapter 2.2 below.

The overall business demography provides the most comprehensive picture of the business dynamics: it includes all enterprises, whether employer enterprises or not. Note that a high share of enterprise births are self-employed enterprises (sole proprietorships). However, these are out of scope in the employer and economic enterprise demography. The reason why the concepts of employer business demography is applied is the improvement of international comparability. Not all countries have data on sole proprietorships available. Thus, excluding sole proprietorships from the scope of business demography increases international comparability.

The reason for the concept of economic enterprise demography is more of conceptual nature. In case that an entrepreneur creates a business and decides to choose the legal form of a limited liability company (or any other form of an incorporated business), he or she himself would be regarded as an employee, rather than as self-employed. Thus, the threshold of two employees ensures comparability with the employer business demography concept. However, an owner-manager of a limited liability company has also the characteristics of an entrepreneur and thus should be counted as a self-employed person in statistics, even if the person is regarded as an employee in administrative registers. If official statistics would treat them as self-employed, the distinction between employer business demography and economic enterprise demography would become obsolete. For more information on the concepts of entrepreneurs, see chapter 4.4.

The international data collections by Eurostat (EU and EFTA countries) and the OECD cover the overall business demography and employer business demography.

2.1.3 Statistical unit

In business statistics, three main statistical units are applied: enterprises, establishments and local units. Business demography statistics can thus also be based on any of these three units. However, in the European Union the main statistical unit is the enterprise and business demography statistics is also based on this unit.

It is evident that the demographic events are different and that the results of the demography data will differ depending on the statistical unit applied.

In case of enterprise based data, an enterprise birth is only counted when no other enterprise is involved in the creation. However, a new establishment of an existing enterprise might be counted as an establishment birth. The same applies for the local unit: an additional new location of an existing enterprise will be counted as a local unit birth. Clearly, in case of single-establishment and single-local unit enterprises the result will be the same irrespective of the unit applied.

A specific complexity in applying the statistical unit is that an enterprise might consist of more than one legal unit. However, this is only the case when these units belong to an enterprise group. The majority of the cases the enterprises will consist of one legal unit only. The majority of enterprise births will also consist of just one legal unit. However, it could also be the case that two or more legal units are created that together form an enterprise birth. Therefore, a newly created legal units have to be checked whether they form an own enterprise or should be allocated to an already existing enterprise. The treatment of legal units that were identified as special purpose entities in business demography is quite straightforward, as they usually do not have neither employment nor turnover. Thus, such legal units are not regarded as active enterprises and are aggregated with the legal unit(s) that build together the enterprise. However, there are cases of legal units, which do have employment and turnover, but are still not regarded as a separate enterprise: such cases could be vertical integration. A further kind of cases could be that the production factors are separated in legal units (e.g. one legal unit cover the building and the machines, another one the employment, etc.). Also in such cases, these separate legal units do not form an enterprise.

International data collections by Eurostat (EU and EFTA countries) and the OECD so far cover business demography data based on the statistical unit enterprise.

2.1.4 Regional business demography

Very much connected with the choice of the statistical unit is the issue of regional business demography statistics. Usually, regional statistics should not just be a view of a national statistic by geography. The concepts should be adapted to focus on regional aspects. This clearly applies for the statistical unit used. The enterprise, which can have locations in more than one region, is thus not the ideal unit for regional statistics. This holds also true for business demography statistics, even if most of the enterprise births are units that are very small and are active in just one region. However, this is not the case for the whole business population and also not for enterprise deaths. The indicators “birth rate” and “death rate” by regions are biased by the use of the statistical unit enterprise as enterprises are counted in that region where they have their headquarters or where they have their main production.

From a regional point of view, the interest is what has happened in the region. In that sense job creations by an enterprise birth is equally important as job creations by additional establishments or local units in that region that are created by already existing enterprises. Thus, the adequate unit for regional business demography is the establishment or the local unit, rather than the enterprise.

Despite of the fact that enterprises are conceptually not the ideal units for regional business demography, international data collections are based on enterprises and are thus national business demography just subdivided by regions. The totals over the regions are the same as the national totals. One reason why enterprises are applied is the fact that due to data availability at the level of local units or establishments the use of these units for regional business demography might not be possible.

2.1.5 Sub-annual business demography

There is principally no conceptual difference between annual and sub-annual business demography statistics. However, data availability might be different and this limits the implementation of the concepts of business demography that are applied for annual data in case of sub-annual demography data. For example, it might be the case that a unit seems to be dormant and information is missing whether the unit is just temporary inactive and might be reactivated in some near future or whether the unit is really dead.

Therefore, for sub-annual business demography statistics less strict definitions might apply and the database is more based on direct administrative sources. Instead of real enterprise births just registration information from company register or other administrative registers will be used. Insofar, sub-annual business demography data are more likely proxies and short-term indicators for business dynamics, rather than conceptually business demography. In considering these weaknesses, sub-annual data are normally shown in the form of time series indicators, rather than in their absolute values.

2.1.6 High-growth enterprises and gazelles

The creation of jobs through new enterprises on the one hand and through fast growing and already existing enterprises are two main anchors of entrepreneurship statistics. High-growth enterprises are defined as those enterprises that have an increase in turnover or employment over a certain threshold and over a certain period of years. It is no relevant whether these enterprises were recently created or not.

A sub-population of high-growth enterprises are gazelles, which are high-growth enterprises that were born in recent years. Also in case of gazelles, thresholds need to be defined and applied.

Data on high-growth enterprises and gazelles are shown by economic activity, size classes and legal form, similarly to the breakdowns in business demography.

2.2 Key concepts and definitions

2.2.1 Demographic events and continuity rules

Demographic events are events with an impact on the existence of statistical units and the links between them. Those events can be split into

- existential changes (i.e. those involving the emergence or disappearance of combination of production factors) and
- distribution changes (i.e. changes in the distribution of production factors between units).

Existential changes mean the emergence or disappearance of statistical units (births and deaths). Due to existential changes, the number of statistical units increases or decreases. Distribution changes between enterprises may result in a decrease of the number of enterprises, or in an increase. The number of enterprises may also remain the same.

The enterprise is the central statistical unit as all statistical units are defined in terms of the enterprise. They are either a part of the enterprise in terms of economic activity or location (kind-of-activity unit and the local unit, respectively), or by both (establishment), or a combination of enterprises bound together by legal and/or financial links (enterprise group). Demographic events are thus relevant for all statistical units (enterprise group, enterprise, kind-of-activity unit, establishment, local unit).

The following table provides an overview of the typology of demographic changes for the unit enterprise:

Event	Real, observable world		Business register	
	Number of enterprises before the event	Number of enterprises after the event	Number of creations	Number of deletions
Birth	-	1	1	-
Death	1	-	-	1
Change of ownership	1	1	-	-
Merger	n	1	1	n
Take-over	n	1	-	n-1
Break-up	1	n	n	1
Split-off	1	n	n-1	-
Creation of a joint venture	n	n+1	1	-
Cessation of a joint venture	n	n-1	-	1
Restructuring within an enterprise	1	1	-	-
Restructuring within an enterprise group	n	n	0 or more	0 or more
Change of group	1	1	-	-
Complex restructuring	n	n	0 or more	0 or more

The explanation of the resulting unit structure caused by the above listed demographic events are as follows:

Birth: Independent event affecting only one enterprise and involves the creation of a new combination of factors of production.

Death: Independent event affecting only one enterprise and involves the dissolution of a combination of factors of production.

Reactivation: Enterprise becoming dormant for a period of less than two years and then recommence activity in a way that complies with the definition of continuity.

Change of ownership: Restructuring of the legal basis of an enterprise. The enterprise remains live and active. No impact on the demographic variables.

Merger: Opposite of a break-up. Consolidation of production factors of two or more enterprises into a new enterprise. There is no continuity or survival, but the closure of the previous enterprises is not considered as deaths. Similarly, the new enterprise is not considered as a birth.

Take-over: Opposite of a split-off. Enterprise taken over are not considered to be deaths. The enterprise that has taken over the other ones survives.

Split-off: Similar to a break-up. Original enterprise does survive. There is no death, but one or more new enterprises are created which are thus not considered as newly born enterprises.

Creation of a joint venture: Two or more independent enterprises agree to commit some of their resources to work together on a common projector towards a common goal. None of the independent enterprises has control over the joint venture. Joint ventures may be considered as a newly born enterprise if they involve the creation of new factors of production.

Cessation of a joint venture: Opposite of a creation of a joint venture. Considered as enterprise death if less than half of the employment is transferred to the participating enterprises.

Restructuring within an enterprise: Involves only one enterprise and has no impact on the demographic variables.

Restructuring within an enterprise group: Involves the creation or cessation of one or more enterprise under common ownership. It does not involve a significant change in the total production factors controlled by the group. It does not result into enterprise births and deaths.

Change of group: Similar to the change of ownership. The enterprise remains live and active. No impact on the demographic variables.

Complex restructure: Similar to the re-structure within an enterprise group, but concerns two or more enterprise groups. Depending on the kind of the re-structure, the outcome as concerns the demographic events might be quite different.

The core conceptual basis for the treatment of demographic events in the statistical business register as well as in business demography are the continuity rules, i.e. the conditions for keeping or changing a statistical units' identity number or not. Various aspects of enterprise characteristics might be considered for defining continuity depending on the use of the resulting data. In the statistical business register and in business demography the continuity concept is based on the continuity of the factors of production. If its factors of production are considered to be continued, the enterprise is considered as continued. It is discontinued if its factors of production are discontinued. Factors of production include employment, machines and equipment, land, buildings, management and intangible assets. In practical implementation all these factors can usually not be analysed in detail, thus some more practical rules are needed.

Considering what is available in the statistical business register and the information supplied by administrative or other sources, the following three criteria proved to be very practical:

Controlling legal unit: The controlling legal unit controls the production factors of the enterprise. The continuity of the management of the enterprise may be assumed to be positively correlated with the continuity of the controlling legal unit. The convention is that there is deemed continuity of the enterprise. A change of the controlling legal unit is not sufficient reason in itself to delete an existing enterprise record and create a new one in the statistical business register.

Economic activity: It is assumed that a change in the economic activity does not imply a change in the factor of production, at least not abruptly. The convention is that a change of the principal activity is not sufficient reason in itself to delete an existing enterprise record and create a new one in the statistical business register.

Location: The continuity of the locations where the activities are carried out is closely linked to the continuity of the land and buildings used by the enterprise. In case of an enterprise ceasing its activities at its main location and resuming its activities at another location within the national territory, the answer whether continuity is given or not, is not obvious. If the activities do not move far, the probability that the production factors other than land and buildings are largely continued is high. However, a long-distance move might mean that also other production factors have changed and thus continuity is not given. The general rule for a change of the main location is that there is deemed continuity. Change of main location is not sufficient reason in itself to delete an existing enterprise record and create a new one in the statistical business register.

From the above it can be concluded that a change just in one of these criteria does (generally) not disrupt continuity. If all three criteria are changing then this means certainly a disruption of continuity. The practical rule is that in case of a change in two of the three criteria, continuity might not be assumed. However, this practical rule should be applied with care. In any event, an analysis of the single cases, especially of big enterprises and complex changes is advisable.

The following table summarizes the continuity rules for the enterprises:

Continuity rules for enterprises								
Change of controlling legal unit	No	Yes	No	No	Yes	No	Yes	Yes
Change of principal activity	No	No	No	Yes	No	Yes	Yes	Yes
Change of main location	No	No	Yes	No	Yes	Yes	No	Yes
Continuity of enterprise?	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**No**	**No**	**No**

In addition to the continuity rules for the enterprise as well as the demographic events, also analogous rules for the enterprise group, the local unit and for the establishment are needed. Continuity rules for the unit enterprise group are not dealt with here.

As stated above all statistical units are defined in terms of the enterprise. If an enterprise is created also a local unit and an establishment is created at the same time. When an enterprise is ceased, also the local units and the establishments are ceased. Thus, demographic events resulting from the enterprise level do also have effects for the local units and establishments of that enterprise.

Considering the continuity rules for local units, location, production factors and principal activity comes into play. The production factors are usually measured by the employment of the unit. A practical convention is that the local unit is deemed to be continuing, if 50% or more of the persons employed by the local unit continue to work at or from the same location. A further aspect is the change of the enterprise to which the local unit belongs. A local unit might be transferred from one enterprise to another. If nothing else changes, the local unit is deemed to have continued.

Concerning the location, the general rule is that a local unit loses its continuity, if the local unit moves to another location, irrespective if the other criteria do change or not. However, a move over a short distance without any changes of one of the other criteria (enterprise, principal activity, and employment) the local unit will retain its identity. There is no harmonised definition of what should be understood as “short distance”. A general rule could be to base it on a regional classification that is in use in the particular country: moving out of the region results in a loss of the identity of the local unit. The regions to which this rule might be applied should be defined as the smallest administrative areas, such a municipalities.

The following table summarizes the continuity rules for local unit under the assumption that there is no change of the location of the unit:

Continuity rules for local units without change of location								
Change of enterprise	No	Yes	No	No	Yes	No	Yes	Yes
Change of principal activity	No	No	No	Yes	No	Yes	Yes	Yes
Change of employment	No	No	Yes	No	Yes	Yes	No	Yes
Continuity of the local unit?	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**No**	**No**	**No**

For the discussion of the demographic events for local units, it is logical to distinguish between demographic events below the enterprise level and above the enterprise level. In the first case the enterprise itself is unchanged, only changes at the level of the local unit are given. In the second case also demographic events affecting the enterprise itself are given. These demographic events affect also the changes at the level of the local units: a birth of an enterprise entails a birth of a local unit; a death of an enterprise means also a death of the local units(s).

In all the other kinds of demographic events of the enterprises (merger, takeover, break-up, split-off, joint venture restructuring) the local units will be transferred to the enterprise that exists after the demographic event. The local units will thus continue its identity, unless other changes do also happen.

However, it could also be the case that in the course of the demographic event, the entrepreneur decides to close one or more of the local units or to create one or more new ones.

The continuity rules and the treatment of the demographic events at the enterprise level for the establishments are the same as for the local units and are thus not further dealt with here.

2.2.2 Coverage

Business demography statistics aims at providing a picture of the development of the business population by data on newly created businesses and the number of jobs created through these new businesses on the one hand and by data on business closures and the jobs that are lost by these closures on the other. Thus, business demography should include all market-oriented businesses of all economic activities. However, usually the agricultural sector is not included. Business demography focusses on market producers, and thus government institutions as well as non-profit institutions are not included.

The delineation of the business demography statistics can best be defined by using the classification of institutional sectors according to SNA 2008/ESA 2010. Included in business demography are the sectors:

- S 11 Non-financial corporation
- S.12 Financial corporations
- S 14.1 Household sector: Employers
- S 14.2 Household sector: Own account workers

For details on the sector classifications and its application see SNA 2008 and ESA 2010 as well as the SBR Guidelines. It should be noted that the subdivision of the S.14 Household sector is only given in the ESA 2010.

By defining the coverage by referencing to the sector classification, additional coverage definitions by economic activities are not needed, except when it should be recorded that certain economic activities should explicitly be excluded (such as ISIC section A Agriculture, forestry and fishing). According to European obligations, the classification of institutional sectors is incorporated into the statistical business registers and thus all institutional units in the register are coded by this classification.

2.2.3 Population of active enterprises

The starting point for the derivation of the business demography events is the definition of the population of active enterprises. It comprises all enterprises that were active in the whole year t or in a part of year t. Activity is defined as producing goods and services for the market and thus having employment (at least one self-employed) or turnover. In case of the employer business and economic business demography the criteria of being active is just measured by the number of employees, which in case of employer enterprises should be at least one and in case of economic enterprises at least two. For the employment criterion, the data should be on an annual average basis (see below 2.3.2).

The number of enterprises in business demography is higher than those that were active at the end of the year as also enterprises are counted as being active in year t, even when they have ceased activity before the end of the year.

2.2.4 Enterprise births

Entries of businesses, i.e. those businesses that are present in a given period but were not present in the previous period, may be easy to identify. However, only a part of the entries are viewed as enterprise births. The main criteria of the definition is that in the process of birth no other enterprise is involved: a birth amounts to the creation of a combination of production factors, with the restriction that no other enterprises are involved in the event. As shown in the previous chapter, enterprises emerging from mergers, split-offs and other demographic events are not viewed a newly born enterprises.

Definition of enterprise births:

A birth amounts to the creation of a combination of production factors with the restriction that no other enterprises are involved in the event: Births do not include entries into the populations due to mergers, break-ups, split-offs or restructuring of a set of enterprises. It does not include entries into a sub-population resulting only from a change of activity.

A birth occurs when an enterprise starts from scratch and actually starts activity. An enterprise creation can be considered an enterprise birth if new production factors, in particular new jobs, are created. If a dormant unit is reactivated within two years, this event is not considered a birth.

The definitions aim as producing data on the creation of new enterprises that have started from scratch and that have actually started activity. An enterprise creation can be considered as an enterprise birth if new production factors, in particular new jobs are created.

Included in the enterprise birth data are enterprises started by a person who previously performed the same activity, but as an employee. Also included are newly born national of foreign subsidiaries, if

- they are real enterprises with autonomy of decision making, and
- new production factors are created, rather than transferred from another unit.

Excluded from the number of enterprise birth should be:

- Enterprises that are created by merging production factors or by splitting them into two (or more) enterprises (break-ups, mergers, split-offs, restructuring),
- Newly created enterprises that simply take over the activity of a previously created enterprise (take-over),
- Any creations of additional legal units/enterprises solely for the purpose of providing a single production factor (e.g. the real estate or personnel) or an ancillary activity for an existing enterprise,
- An enterprise that is newly registered when an existing enterprise changes legal form,
- Reactivated enterprises if they restart activity within 2 calendar years.

However, as three kinds of business demography concepts are distinguished, also the enterprise birth definition needs to be adapted: only in the case of the overall business demography, the focus of the definition is on the fact of creating a combination of production factors in order to qualify a new enterprise entry as a birth. In case of the employer business demography and the economic business demography the decisive criterion is the number of employees. An employer enterprise birth is given when the enterprise has at least one employee and in case of economic enterprise two employees. This can happen in the year of the actual creation or in a later period. See the following table:

Enterprise births according to the business demography concepts	
Enterprise births	Births of all enterprises, regardless of whether they are employers or not. No general threshold is applied to the size of the enterprise in terms of employment or any other characteristics.
Employer enterprise births	Births of enterprises with at least one employee. This population consists of births that have at least one employee in the birth year and of enterprises that existed before the year in consideration, but were below the threshold of one employee (entry by growth)
Economic enterprise births	Births of enterprises with at least two employees. This population consists of enterprise births that have at least two employees in the birth year and of enterprises that existed before the year in consideration, but were below the threshold of two employees (entry by growth).

The same unit may be recorded as a birth in more than one of these populations: a newly created enterprise may be counted as a birth when the economic activity started, irrespective of whether it

has employees or not. In case of one employee from the beginning, this unit is part of the overall birth population as well as of the population of employer enterprise births. Similarly, in case of two or more employees the unit would also be part of the economic enterprises births. However, these events might not happen all in the same period (year). An enterprise may start with no employees and might engage its first employee in a later than the creation period. Thus, the same unit will be part of the employer enterprise births in a later period (year).

The following table lists the possible cases when an enterprise will be recorded as a birth in the three different business demography concepts:

Cases of enterprise births	**Business demography**	**Employer business demography**	**Economic business demography**
Enterprise A starts its activities in t only with an self-employed person	Birth in t	Out of scope	Out of scope
Enterprise B starts its activities in t with 1 employee from the beginning	Birth in t	Birth in t	Out of scope
Enterprise C starts its activities in t with 2 employees from the beginning	Birth in t	Birth in t	Birth in t
Enterprise A engages in t+1 one employee	Out of scope	Birth in t+1	Out of scope
Enterprise A engages in t+2 one further employee	Out of scope	Out of scope	Birth in t+2
Enterprise B engages in t+1 one further employee	Out of scope	Out of scope	Birth in t+1

A question related to the definition of enterprise births is the question when such a birth happens and how it should be calculated. Again, in case of the employer and economic enterprise the solution is conceptual easy: such units should be recorded as births in the year when the employees' threshold has been reached. In case of business demography, several definition to recognise a new enterprise as births could be possible: an enterprise may be viewed as starting its activities already when investments are made, thus even before any production activities and sales on the market are done. Another possibility could be to record an enterprise as born when the enterprise has started its production activities. An indicator for that would be turnover/sales. It is also an indication that production activity is going to start when the enterprise has engaged one or more employees. Usually, data on turnover and employment are more likely to be available as information on preceding investment expenditures.

2.2.5 Enterprise survival

The second focus of business demography is on the survival of newly born enterprises over the subsequent years. This type of data is based on the analysis of a cohort of enterprise births in year t. It thus allows to follow up the development of newly born enterprises over the years: how many have survived and what is their economic development in form of employment and/or turnover. Usually, the time horizon for that type of analysis is three to five years. For each year, a survival rate is calculated. The survival rates decrease over the years and after five years only about half of the enterprises may still be active or in other words, half of the enterprises have ceased their activity. Of course, business demography statistics cannot provide any data on the reasons of the cessation of activity. For this goal additional information might be needed which may only be supplied directly from the enterprise.

The following box contains the definition of enterprise survival:

Definition of enterprise survival:
Survival of an enterprise occurs if an enterprise is active in terms of employment and/or turnover in the year of birth and the following year(s). Two types of survival can be distinguished:

1.) An enterprise born in year t is considered to have survived in year t+1 if it is active in terms of turnover and/or employment in any part of year t+1 (= survival without changes).

2.) An enterprise is also considered to have survived if the linked legal unit(s) have ceased to be active, but their activity has been taken over by a new legal unit set up specifically to take over the factors of production of that enterprise (= survival by take-over):

The definition of survivals excludes cases where enterprises merge or are taken over by an existing enterprise in year t. In these cases the continuation of the enterprise involves an enterprise established before year t+1and therefore the enterprise is not considered to have survived.

The definition also implies that the enterprise that takes over the factors of production is a new enterprise, i. e. an enterprise that commences activity in the year of the take-over and that is not a reactivation. Lastly, it should be noted that for consistency reasons the definitions of births and survival should be followed in equal measure.

The following table summarizes the definitions of survival for the three kinds of business demography:

Enterprise survival according to the business demography concepts:	
Enterprise demography	An enterprise born in year t has survived in year t+1 if is active in terms of turnover or employment in year t+1.
Employer enterprise demography	An employer enterprise born in year t has survived in year t+1 if it has at least one employee in year t+1.
Economic enterprise demography	An economic enterprise born in year t has survived in year t+1 if it has at least two employees in year t+1.

These rules are illustrated below:

Cases of enterprise survival	**Business demography**	**Employer business demography**	**Economic business demography**
Enterprise A continues its activities in t+1 only with an self-employed person	Survival in t+1	Out of scope	Out of scope
Enterprise B continues its activities in t+1 with 1 employee	Survival in t+1	Survival in t+1	Out of scope
Enterprise C continues its activities in t+1 with 2 employees	Survival in t+1	Survival in t+1t	Survival in t+1

The identification of units that have survived should always be done on a year-to-year basis: an enterprise born in year t has survived in year t+2 only if it has also survived in year t+1. The methodology for the identification of survivals is the same for all consecutive years starting from the year of birth. The criteria of survival are of course different depending on the kind of business demography that is applied: for business demography it is the criteria of being active in the sense of turnover or employment, for employer business demography it is the hiring of at least one employees and for the economic business demography it is the hiring of at least two employees.

2.2.6 Enterprise deaths

Like enterprise births, enterprise deaths, are also identified based on the whole business population. Data on death enterprises thus refer to enterprises that were born in any year, not necessarily just in the preceding years. The definition of enterprise death is as follows:

Definition of enterprise deaths:
A death amounts to the dissolution of a combination of production factors with the restriction that no other enterprises are involved in the event. Deaths do not include exists from the population due to mergers, take-overs, break-ups or restructuring of a set of enterprises. It does not include exits from a sub-population resulting only from a change of activity.

According to the three different concepts of business demography, the definitions of deaths are as follows:

Enterprise deaths according to the business demography concepts:	
Enterprise demography	Enterprise deaths cover all enterprises, regardless of whether they are employers or not. No general threshold is applied to the size of the enterprise in terms of employment or any other characteristics.
Employer enterprise demography	An employer enterprise death occurs either as an enterprise death with at least one employee in the year of death or as an exit by decline, moving below the threshold of one employee.
Economic enterprise demography	An economic enterprise death occurs either as an enterprise death with at least two employees in the year of death or as an exit by decline, moving below the threshold of two employees.

Like in the case of births also in the case of deaths, the same unit may be recorded as a death in more than one of these populations. For instance, an enterprise may move below the threshold of one employee in year t+1 and thus will be counted as employer enterprise death, and may cease all its activity in t+2, which means an enterprise death in t+2. The following table provides some examples:

Cases of enterprise deaths	**Business demography**	**Employer business demography**	**Economic business demography**
Enterprise A ceases its activities in t+1	Death in t+1	Out of scope	Out of scope
Enterprise B with one employee ceases its activities in t+1	Death in t+1	Death in t+1	Out of scope
Enterprise C with two employees ceases its activities in t+1	Death in t+1	Death in t+1	Death in t+1
Enterprise B reduces its employment from one employee to zero employee in t+1	Survival in t+1	Death in t+1	Out of scope
Enterprise C reduces its employment from 2 to one employees in t+1	Survival in t+1	Survival in t+1	Death in t+1
Enterprise C ceases its activities in year t+2	Death in year t+2	Death in year t+2	Out of scope

Like in the case of births and survivals, the definitions should be applied in the analogous ways so that conceptual consistency between the birth, survival and death data will be achieved. In case of deaths, this means that the following enterprise closures should not be counted as enterprise deaths:

- Enterprises that close down due to merging or breaking-up of production factors,
- Enterprises whose activity is taken over by another enterprise,
- Enterprises that are deleted due to a change of legal form (for example. sole proprietor changes its legal form to a limited company),
- Reactivated enterprises if they restart activity within 2 calendar years.

Enterprise deaths should be counted in that period where the enterprise has ceased its production activities. This means that some employment or turnover due to the ongoing liquidation process may still be given. Also that legal unit(s) may still exist for some further periods and are still recorded as (active) units in administrative databases.

The recording of enterprise deaths are in practice more difficult than the recording of enterprise births. One reason is that information on the data of cessation may not be available in a timely manner from administrative sources or may reflect only administrative deaths. Thus, additional data may have to be used, such as information on turnover or employment. However, this information may also not be available in a timely manner, especially turnover data from taxation. A further difficulty is that an enterprise cessation should not be counted as an enterprise death, if the enterprise re-starts its activity within two years. It is thus logical that death data might be revised or updated regularly, which is usually not the case with birth data.

2.3 Variables and characteristics of business demography

This sub-chapter discusses those variables and characteristics that form the core of business demography statistics. In case of micro-data linking of business demography units with other databases additional variables and characteristics will emerge. Such variables could be engagement in international trade, in R&D, etc. (see chapter 4). These variables are not discussed in this sub-chapter.

2.3.1 Activity classification

For reasons of international comparability, the ISIC Rev. 4 activity classification should be used in business demography statistics. In most countries, the ISIC Rev. 4 classification is applied in the statistical business registers and thus activity codes are available. In Europe, a more detailed, but fully compatible activity classification (NACE Rev. 2) is obligatory by European Regulation. All issues concerning the application of an activity classification need not to be dealt with here; reference is given to the ISIC and the SBR manual.

As concerns the level of application it is recommended to use detailed classification levels of ISIC Rev. 4 in order to provide a detailed picture of the business dynamics. Of course the level of detail might be influenced by data and confidentiality constraints, so that the lowest level can not be applied or not in all activities. As most of the newly born enterprises are active in service industries, it is recommended that a more detailed level of services part of ISIC might be applied as for the other economic activities. However, any national adaptations of ISIC Rev. 4 for national purposes should be in line with the hierarchical structure of the classification so not to disturb international comparability.

However, there is one specific aspect in the use of the activity code of the units. In order to establish survival time series, any changes in the activity code of a newly born unit would not allow to follow a birth cohort of year t over the subsequent (five) years by economic activities. Of course, changes in the main economic activity take place in reality and are to be incorporated in the statistical business register according to the appropriate rules. However for business demography the stability of the activity code is needed, namely that code that was valid in the period of the birth. This means that in the longitudinal database that needs to be created for the purpose of business demography statistics, the activity code of the year of birth should be included.

However, in case of high-growth enterprises, their economic activities should be observed at the end of the growth period rather than at the beginning. If there is a change in activity during the growth period, the activity into which an enterprise grew can be considered more relevant than the activity out of which it started to grow.

2.3.2 Employment

The most important variable in business demography is employment: self-employed and employees. Employment data are shown for the whole population of active businesses, employment created by newly born units, employment losses by closing of businesses and development of employment of newly born units over subsequent years. Employment data are also needed to subdivide the population into employer businesses and economic businesses on the hand and the businesses that have no employees on the other. Employment is also the main size class criteria (see 2.3.4).

The definition of employment in business demography should be the same as for other business statistics. SNA 2008/ESA 2010 provide a basic definition:

Definition of employees:

Employees are defined as all persons who, by agreement, work for another resident institutional unit and receive compensation in the form of wages, salaries, fees, gratuities, piecework pay or remuneration in kind. An employer-employee relationship exists when there is an agreement, which may be formal or informal, between an enterprise and a person, normally entered into voluntarily by both parties, whereby the person works for the enterprise in return or remuneration in cash or in kind.

Employees include part-time workers, seasonal workers and persons on strike or on short-term leave, but excludes persons on long-term leave. Employees do not include voluntary workers.

A person is considered to be wage or salary earner of a particular unit if he or she receives a wage or salary from the unit regardless of where the work is done (inside or outside the production unit). A worker form a temporary employment agency is considered to be an employee of the temporary employment agency and not of the unit (customer) in which he or she is working.

Definition of self-employment:

Self-employed persons are defined as persons who are the sole owner, or joint owners, of the unincorporated enterprises in which they work, excluding those unincorporated enterprises that are classified as quasi-corporations. Self-employed persons are classified here if they are not also in paid employment which constitutes their principal activity: in the latter case they are under employees. Self-employed persons also include the following categories: unpaid family workers, outworkers and worker engaged in production undertaken entirely for their own final consumption or capital formation, either individually or collectively.

From an economic point of view, a person managing and owning a limited liability company (or any other form of an incorporated business) should be counted as an employer rather than as an employee. However, in practice it might not be easy to distinguish. Also in the current international definitions in business statistics, the definition of an employee includes paid working proprietors.

While in economic statistics and in national accounts employment data in full time equivalents or even more in time measures (hours worked) are the preferred measurement concepts, in business demography data head counts are used. In economic statistics labour is mainly seen from the resources' inputs into the production process, in business demography the main interest is in the number of jobs, irrespective whether these are full time jobs or not.

Employment should be calculated on an annual average basis. Thus, in principle more than one employment information per year for each businesses should be available to calculate an annual average. Ideally, monthly employment data should be available.

The annual average is to be calculated in case of monthly data by dividing the aggregated total employment by 12 and in case of quarterly data by 4. Of course, if there is only one employment data available, for instance for the end of the year, annual averages cannot really be calculated; additional estimates might be necessary. The conceptual question is whether annual averages should be applied for all employment data with or without considering the real length of the operating period (employment of the population of active business, employment in newly born units, employment of survived units, employment of closed units, and for the determination of the size class allocation).

In case that an enterprise is not active over the whole year, born in some point in time during the year, or closed in some time during the year, the conceptual question is whether annual averages should be applied also in these cases or averages that only take into account the length of the operating period. The annual average is the main employment concept in economic statistics as jobs are viewed as a

labour input indicator to the economic production process. In business demography statistics – where the main interest is in the development of the number of jobs - it could be argued that an average over the operating period which could be less than a year is more adequate. For all businesses which were operating the whole year there is clearly no difference between the results of these methods.

However, for the units that have a shorter operating period than the whole year the average over the operating period results in higher employment figures than in case of an annual average. For the whole population of active businesses the total effect will be quite small, as most of the enterprises operate over the whole year.

Let us assume that an enterprise was born in October of year t and has 4 employees in each of the remaining months of year t. The operating period of that unit is thus three months, so the average method employment based on the operating period concept would be 4, whereas in case of the annual average the average employment would be 1. The number of employment by end of year t (compared to year t-1) increased by 4. The average number of employment of year t (compared to year t-1) increased by just 1. So, the labour input into the production process of year t increased by one employee. The newly born enterprise has created four jobs, but not for the whole year. However, if it is assumed that the 4 jobs at the end of year t would be sustainable, the business demography message is that employment has increased by 4.

According to the current international concepts, the calculation of the averages should be based on the operating period for birth and death data. This will result into a measure of number of jobs created by newly born enterprises and number of jobs lost by cessation of enterprises.

2.3.3 Turnover

Turnover is one of the output measures of economic activities. Strictly speaking, turnover (sales to the market) is a variable that can be observed for the unit enterprise only. In case of establishments, the output measure is the total output, which includes also deliveries between the establishments of the enterprise. In the case of business demography statistics, turnover plays not a central role. In the obligatory data delivery programme of the European Union no turnover data are requested. One of the reasons is that turnover data – especially when taken from administrative sources may have long time lags and in general data availability in a timely manner is not so common in many countries. Another reason is of course the focus on job creations related to enterprise births in business demography statistics. A further reason is that due to the differing economic development of the countries, the different currencies and purchasing power parities, the monetary data on turnover are not easy to compare internationally.

Definition of turnover:

Turnover comprises the totals invoiced by the observation unit during the reference period, and it corresponds to market sales of goods and services supplied to third parties.

Turnover includes all duties and taxes on the goods or services invoiced by the unit with the exception of the VAT invoiced by the unit vis-a vis its customer, and other deductible taxes directly linked to turnover.

It also includes all other charges (transport, packages, etc.) passed on to the customer. Reduction in prices, rebates and discounts as well as the value of returned packaging must be deducted.

2.3.4 Size classes

In principle, there are two indicator for size of the statistical units: the turnover and the employment. As said above turnover data do not play a significant role in business demography statistics. This is also the case of turnover as an indicator for size of an enterprise.

The central size indicator is based on the number of employees. In the European Union five employee-based size classes are applied:

0 employees

From 1 to 4 employees

From 5 to 9 employees

10 employees and more

As the size classes are based on the number of employees only and not on total employment, a size class of zero employees is needed. As the newly born enterprises are usually quite small, the size-class structure for business demography statistics focuses on the smaller number of employees. The biggest size-class refers to 10 or more employees. Of course, the size-class structure may be adapted for national purposes. For instance by additional size-class elements for enterprises with more than 10 employees may be added. However, any national adaptation should be done in such a way that it can be transformed into the international one.

Data by size classes are not only important for the population of active enterprises but also for business dynamics of births and deaths. An enterprise may grow after its birth and may in reality belong to another size class in the surviving year(s) than in the birth year. However, as the interesting indicator is the growth of employment of a population of newly born enterprises, the surviving enterprise should stay in the initial size class. This is analogous to the concept that the activity code of surviving enterprises should still be the code of the year of birth.

As the size class allocation is determined on the basis of average employment data, conventions are needed for the boundaries:

Employee size class	**Boundaries in number of employees**
0 employees	0 employees
1 to 4 employees	>0 to > 4.5
5 – 9 employees	≥4.5 to >9.5
10 or more employees	≥9.5 employees

2.3.5 Legal forms

A further variable to describe the units is the legal form of the enterprise. Legal form can only be applied for the unit enterprise and not for the other statistical units. However, strictly speaking even in case of an enterprise the legal form might not be evident. This is the case when an enterprise consists of more than one legal unit (and these legal units have different legal forms). For such cases, there is currently no agreed international concept. One solution might be to classify the enterprise according to the legal form of that legal unit which is the main producing unit. Another option could be to use the legal form of that legal unit which has most of employment or value added.

The choice of a legal form when creating an enterprise depends on various criteria. The most relevant are certainly the kind of liability that the entrepreneur will be faced as well as taxation or other administrative criteria. These criteria may change over time and thus the choice of a specific legal form may change over time. Furthermore, there might be political or taxation incentives in favour of a specific legal form.

Another important issue is that the legal forms may differ considerably between the countries. For comparability reason it is thus advisable to use a classification of legal forms that is quite aggregated. For instance, in the European Union only three kinds of legal forms are distinguished:

- Personally owned and no limit to personal liability
- Private of publicly quoted joint stock companies with limited liability for those owning shares

- Personally owned limited and unlimited liability partnerships. Included are also other legal forms, such as cooperatives, associations, etc.

The three forms above are not relating to concrete and specific kinds of legal form, but are rather constructs based on characteristics of the desired legal forms. The mapping of the national legal forms to the international classification has to be done by the national statistical office itself.

2.4 Key concepts and definitions of high-growth enterprises and gazelles

2.4.1 High-growth enterprises

The concept of high-growth enterprises is not directly related to business demography concepts. The identification of high-growth enterprises is based on the whole population of active enterprises. High-growth enterprises are defined as those enterprises that satisfy some predetermined threshold that distinguishes them as high-growth over a certain period and that have to be above a certain size to mitigate any small enterprise growth bias. Growth can be measured both in terms of employment (number of employees) and in terms of turnover. However, in practical implementation and because of the difficulties in interpreting monetary turnover data over the countries, international data collections use the number of employees to identify the high-growth enterprises.

Definition of high-growth enterprises:

High-growth enterprises are enterprises with average annualised growth greater than 10% per annum, over a three-year period. Growth can be measured by the number of employees or by turnover.

If growth in the number of employees or turnover was due to mergers or take-overs, the enterprise in question should not be considered a high-growth enterprise.

In the Eurostat/OECD business demography manual of 2007 the threshold of the growth rate was 20%. Based on results of pilot studies the threshold in international data reporting was reduced to 10%. The main reason for this reduction was clearly that the number of high-growth enterprises is relatively small. This results also in confidentiality problems.

This definition thus refers to an annual growth rate of 10% or more over a period of three years. In order not to distort the resulting data set by enterprises that have a small number of employees a size threshold is used. An enterprise that grew from one to two employees would be qualified as high-growth enterprise. In the European implementation, this size threshold is currently 10 employees at the beginning of the growth period.

It is also recommended that enterprises born in year t should not be included in the three years period under investigation as it could be that in the year of birth the enterprise existed only for a period less than a year and would thus have a smaller annual average number of employees. Thus, for this unit a higher growth would result.

According to the definition, it is not necessary that an annual growth of 10% or more be given for each of the years of the three-year period: an average annualised growth is sufficient. As stated in 2.3.1 the activity code of the final period should be used for stratifying the high-growth enterprises, as it can be assumed that this activity might be the one, which served the growth better than an activity at the beginning of the observation period.

Calculating high-growth data on an annual basis an enterprise might be included in the population of high-growth enterprises in several years. From a policy point of view, it might thus be interested to split the high-growth enterprises newly entered in this population and those that were high-growth already in earlier observation periods.

As indicated in the definition growth by mergers and takeovers should not be considered as a real growth. This means that of course for the calculation of high-growth enterprises an appropriate longitudinal database of enterprises is available that corresponds to the business demography concepts.

2.4.2 Gazelles

Gazelles are a subset of high-growth enterprises, which are up to five years old.

Definition of gazelles:

Gazelles are enterprises up to 5 years old with average annualised growth greater than 10% per annum, over a three-year period. Growth can be measured by the number of employees or by turnover.

The basic concepts of gazelles are the same as of the high-growth enterprises. The only difference in scope and data reporting is that only the populations of newly born enterprises (not older than 5 years) are considered rather than the entire populations of active enterprises.

Chapter 3: Requirements of SBR system to support the production of demography statistics

For the purpose of fulfilling its role of supporting the production of business demography statistics, this chapter, in essence, aims at building the bridge between the concepts of business demography and the necessary infrastructure in the statistical business registers to support the production of business demography statistics. Chapter 3 focuses on how to develop Statistical Business Registers (SBR) to facilitate the production of business demography statistics. It describes, recommends and advises on refinements to the SBR UNECE guidelines.

The major assumption in chapter 3 is that a SBR is available, or is in development, and that this SBR follows to the greatest extent as possible the UNECE SBR guidelines. Some of the SBR component, decision rules and concepts, be it the live register, snapshots and historical frame, logs, statistical units, demographic events and continuity rules, SBR maintenance and update, are directly linked to the creation of a database to support the production of business demography statistics.

While a SBR that follows the guidelines holds the needed information to support the production of business demography statistics, its structure does not necessarily allow to readily and efficiently generate business demography statistics. A SBR developed primarily to serve as survey frames is not likely organized in a way to provide a longitudinal image of the statistical units, as recording and keeping track of all historical events may not be a priority.

Basically, a SBR supporting the production of business demography statistics must 1) be historically adjusted and updated (retroactive updates to business characteristics – classification, size and statuses), and 2) allow for easily tracking of businesses through time (continuity).

Chapter 3 first highlight the role of the Statistical Business Register (SBR) as related to the production of business demography statistics. Subsection 3.2 presents the concept of longitudinal business information in the context of the SBR. The characteristics of a longitudinal business database are detailed in sub-section 3.3. The following section add elements on data sources, coverage, statistical units and population that were not covered in Chapter 2. Sub-section 3.5 presents the longitudinal database system requirements, including maintenance and update schedule, corrections and changes, business rules, governance and IT considerations. The chapter's last section present data model examples from countries that have already developed a SBR based longitudinal business database.

3.1 The SBR and the Production of Business Demography Statistics

The UNECE Guidelines on SBR identifies eight roles for the SBR (Section 2.1). The Guidelines on Statistical Business Registers recognized that SBR should be the backbone for the production of business demography statistics at the NSI's. However, the current SBR guidelines are not directly fitted to produce business demography statistics.

3.1.1 SBR Roles – Production of economic statistics

The UNECE Guidelines on SBR identifies eight roles for the SBR (Section 2.1). The Guidelines on Statistical Business Registers recognized that SBR should be the backbone for the production of business demography statistics at the NSI's.

Role number 5 (SBR based statistics – Section 2.6) states that the SBR can "play a role in providing more information at less cost" on its own or in combination with other administrative registers or statistical data sources. Paragraph 2.51 of the UNECE SBR Guidelines states that the SBR is "*…an ideal source in terms of coverage and costs for the production and dissemination of enterprise business demographic statistics."*

The SBR holds consistent and comparable information at the country level and internationally, that can satisfy the requirements for the production of enterprise demography statistics. Indeed, all or most of the information of interest are already found in the SBR, and the SBR offers record linkage facility with other statistical and administrative data files.

However, a SBR developed to serve primarily as a survey frame may not include all of the features and needed information to easily and readily generate coherent business demography statistics.

3.1.2 SBR Shortcomings related to Business Demography

The Guidelines on Statistical Business Register provides direction on dealing with demographic events (Chapter 7.5). In the SBR, the outcome of demographic events (birth, death or continuity) is guided by the SBR main roles, which are to serve as a frame and to support NSIs survey programs. The continuity of statistical units in the context of the production and maintenance of survey frame does not always align to the definition of continuity for the purpose of business demography statistics.

According to the SBR Guidelines, a statistical unit is deemed to continue through a demographic event if two out of three of its production factors continue (*controlling legal unit, economic activity and location*). However, the guidelines stipulate that a change in ownership typically results in the death of a unit and in the birth of another one in SBRs that are maintained from an administrative data source; and even if two of its production factors continue (*economic activity and location*). The same can apply in cases of demographic events (mergers and takeovers, break-ups and split-offs) where new identity numbers – statistical identifiers – are created for emerging units following the event.

Some of the statistical units ceased and created in a SBR supporting survey frame and programs shall rather translate into business continuity in the context of business demography statistics. As such, the population of statistical units must be cleaned of these "false" or "spurious" births and deaths. This is an important step to the production of unbiased business demography statistics from the SBR.

It is indeed essential to reassess the outcome of demographic events in the SBR population. Reassessment of outcome of demographic events to identify continuity where the SBR recorded births and deaths can be achieved through record linkage technics and considering continuity of the SBR production factors. The newly established linkages between a ceased (death) unit and a created (birth) statistical unit must be recorded in some way. This can be done through a series of tables listing linked units or by assigning a new and common longitudinal identifier to the linked units.

3.2 Longitudinal Business Information and the SBR

A longitudinal[4] representation of statistical units' information from the SBR allows to identify continuous businesses vs. greenfield births and closures, it also permits the production of comparable business demography statistics, over time, through consecutive reference periods, at two non-consecutive points in time, or to follow individual statistical units through time.

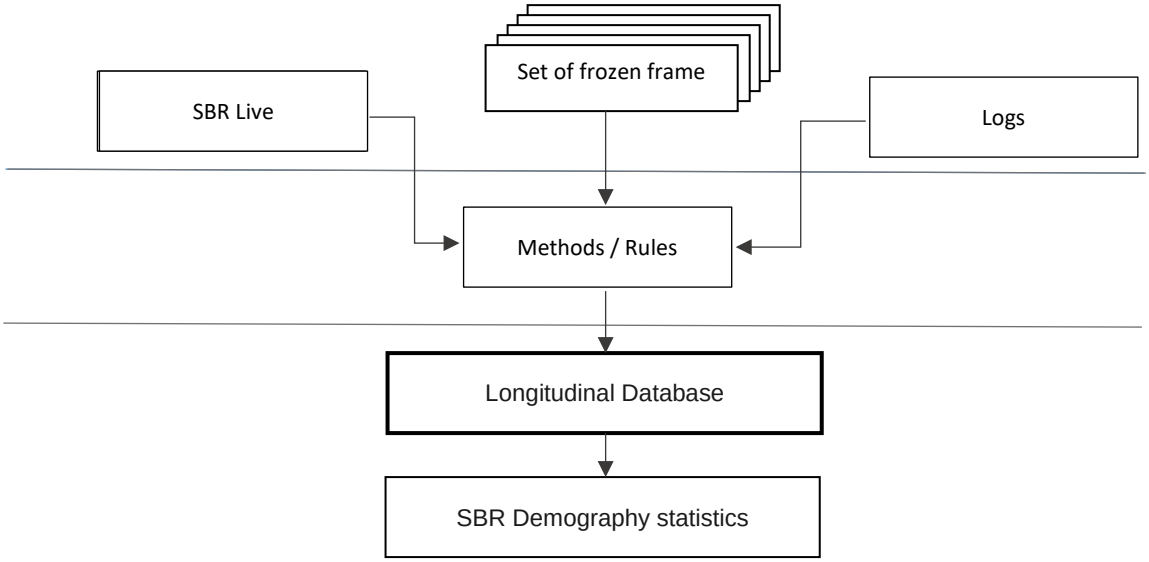
A database in support of demographic statistics has to provide characteristics and size information for statistical units by reference period from the date of creation to time of closure of the businesses.

[4] Longitudinal data refers to data on a single entity collected at different points in time, as opposed to cross-sectional data.

During the course of its life, a business may go through events that modify one or more of its production factors. In the SBR these changes in production factors can translate in businesses birth, death or continuity. They may also reveal fast growing enterprises (gazelles), and allow for the identification of fast growing sectors or regions.

In the context of a SBR, the history of statistical units is created by linking information from a series of tables (live register, snapshots and frozen frames) and logs. Assembling data from multiple tables and creating historical views of statistical units, taking into account impacts of demographic events, require the use of record linkage methods. The longitudinal representation of a single business may link multiple statistical units and as such calls for the creation of a new longitudinal statistical units' identification scheme.

Figure 3.1: SBR based demography statistics



Assembling these different pieces of information into a continuous image of the statistical units[5] is a complex and time consuming task. The process cannot be run each time business demography statistics are to be created. Once assembled the continuous image of the statistical units must be included in a database/table(s) and updated on regular basis.

The new longitudinal database should be designed with sufficient flexibility to allow for the addition of new information, for example at the upcoming SBR redesign cycles, and to accommodate past SBR changes. Due to the regular SBR update cycles, content of the snapshots (variables included), likely differ from the live register. It is expected that newer information may not be available for the earlier years of the period covered by the longitudinal database (LD).

3.2.1 Live BR, snapshots and logs (creation of the LD)

The SBR holds a live register, a series of snapshots and logs. The creation of the longitudinal database calls for the use of all of these SBR data files, along with historical administrative data sources. Statistical units' information included in these data tables need to be linked and organized in order to construct a continuous, from past to current, image of the units.

The live register contains the most up to date image of the SBR population. As stated in paragraph 2.52 of the UNECE SBR guidelines, the live register is refreshed on a continuous basis with the latest

[5] Note that not all snapshots will be required for each and every statistical unit. There will be cases where units kept the same basic information through time.

information available. The live register includes statistical units' size and characteristics information for the most recent reference period. The reference period in the SBR is specific to each administrative data source and to individual statistical units. Size variables (turn-over, employment, assets) may cover a year (fiscal or calendar year), a quarter or a month. As well, most recent period may not correspond between sources. Timing of the update to the classification variables in the SBR also varies between the types of variables and by statistical units.

The series of frozen frames/snapshots consists of past images of the live register. These images are frozen in time. The concept of retroactivity does not exist in the snapshots, and with time, snapshots become out-dated images of the past.

Updates to the live register are recorded into logs, the record keeper of changes in the SBR. For example, they hold new and old values, source of the update, reasons for changes, date of correction and effective date of change[6], to name a few. Information stored in the logs are therefore essential to construct the historical image of the statistical units. The logs entries can serve to overcome the lack of retroactive update of the snapshots. But logs have limitations as well. New information on statistical units for past period, e.g. covering snapshots or reference periods previous to the live register, may not be recorded in the live register, and as such could be missing from the logs.

Therefore, some information required to create a full and complete continuous image of statistical units may not all be available in the SBR (missing from live register, snapshots and logs). Depending on the significance of the missing information in the production of business demography statistics, it may be decided to either accept the information gap, or to revert to other longitudinal data sources, like an administrative data files for example, or to impute missing values based on system rules and accepted assumptions. Decisions to fill or not missing information shall be well documented by NSIs. Alternative data sources and rules and methods to create synthetic data shall also be supported with proper documentation.

Once the longitudinal database is created the SBR snapshots would have served their purpose. At that point, all information needed to update the LD with the new and most recent information can be found in the SBR live register the logs and the administrative data sources. The use of the different sources (SBR live, logs and administrative data sources) is a function of the method, approach and frequency of the LD update and maintenance processes.

3.2.2 Creation of the Longitudinal Database - Past History and Looking in the Future

There is a clear distinction between establishing a longitudinal database initially (history) and the current maintenance of the database (future oriented). While past/historical information are needed to establish a longitudinal database, decisions on the content, the scope and characteristics of the database should be future oriented, mainly driven by the current and future state of the SBR and business demography statistics, and of the availability of input data sources.

Some information included in the SBR live register might not be available in past or previous periods. The information may vary between snapshots due to the addition of new data sources, changes in methods, or following SBR redesign cycles for example. Historical image may be readily available or easily derivable from data included in the SBR, and this information may go several years back, for some variables or types of information. Other data may only be available for a limited number of years, or may show significant changes in content or format at a certain point in time. These

[6] In the context of business demography and the construction of continuous image of statistical units, the preferred date of changes shall be the effective date, i.e. the date at which the change occurred in the "real world", not on the SBR database. If the effective date is not available, then a best approximate date can be used.

inconsistencies in the content or structure of the snapshots may lead to limitations in the construction of the past history.

While the above must be taken into consideration when establishing the longitudinal database, decision on the number of years included in the past should be based on the SBR statistical units' information (population, concepts, methods, characteristics). The major consideration should be the completeness and historical comparability of the population in the BR snapshots. The earliest year should start at a point in time where previous year's population do not compare well with the current nature of the SBR. Alternately, it should be at the point in time where the longitudinal information would show a break in the series that cannot be explained by data itself, but due to methodological or conceptual changes in the SBR.

Another driver about the part of history that must be included in the longitudinal database would be the window supporting the production of flagship business demography statistics (and entrepreneurship indicators) at an acceptable level quality. NSI may set this window to a minimum of 4 years for example.

The availability of historical information for administrative data sources should only be second to the above issue on the selection of the starting date of the longitudinal database. Starting date will vary by administrative data sources, some going back many years while others have become available only recently.

Nonetheless, when first implementing a LD NSI's shall aim at covering at a minimum a period of six years, i.e. current reference period and going back five years. This six-year span allows for the production of five-year survival rates statistics, one of the key measures of business demography. Once established, the longitudinal database time series will continue to grow and shall include plans to add new data sources, for example through supplemental record linkages (other registers).

3.2.3 Identification of continuing businesses

The SBR guidelines discuss the need for record linkage methods in the SBR. The guidelines document defines criteria to link statistical units (e.g. types of matching, name and address standardization, string matching, blocking, etc.). In the guidelines record linkage is primarily for the purpose of updating the live register or to add new sources of information to the SBR. Record linkage allows to associate newly received information (e.g. administrative data) to the SBR and to translate this signal in the creation of a new statistical unit (birth) or in the update to an existing unit (survival/continuity).

Birth, Death or Continuity

As a side note, before a new unit is created on the SBR, the population of the SBR live register should be scanned using a record linkage process to ensure that any new statistical unit is not already represented in the population. This search for unit's continuity will ensure that newly created units are true/real births, in the context of business demography and the longitudinal representation of statistical units.

Units should be match within reference period to find cases of duplications. Units from two consecutive reference periods should also be match. All units matched in the consecutive periods are deemed to continue. Units not linked to another unit in the previous or the following period (non continuity) are probably affected by a demographic event. Whenever possible, information should be retrieved from the logs and input data (administrative data in the case of legal entities) to categorize the demographic event leading to the cessation or creation of the statistical units, along with the date of occurrence. This information will help to better identify potential predecessor or successor to the units, or to set effective date of birth or death of the units.

Skipping this step may lead to duplication of statistical units, the creation of spurious births, or the unduly inactivation of units. These are known to create biases in business demographic statistics. A record linkage method is also essential to the creation of the longitudinal image of the statistical units and the longitudinal business database.

Record Linkage Method

A sound enterprise record linkage methodology is a pre-requisite to the building of a longitudinal database from the SBR. Indeed, information from all SBR tables (e.g. live register, snapshots and frozen frames) must be assembled using record linkage techniques and methods to create the continuous history of the statistical units.

The recommendation in the SBR guidelines manual on record linkage – based on existing identifiers (e.g. from administrative data sources), or on the units' names and address of the units – are applicable to link records for the creation of longitudinal image of statistical units.

The success rate of linkages impact the quality of the longitudinal data and the business demography statistics derived from it. It is advisable to adopt methods, technics and approaches that can improve the record linkage success in the context of a longitudinal database. For example, employer-employee based record linkages may help identifying new relations in the population of statistical units[7]. Employer-employee linkages can help in the identification of continuing units, and reduce the number of false birth and deaths. The use of the phone number or of geographic coordinates could also help identify through linkages. Applying new linkage rules on the statistical units' population will reveal new linkages and wrongly matched records.

Ideally, whenever changes are made to the record linkage methods, SBR tables should be scanned over again to identify previously missed linkages or to break wrongly created linkages. The outcome of the exercise shall lead to a better representation of the longitudinal image of the population of statistical units.

3.2.4 Longitudinal business identifier – tracking units through time

For longitudinal studies units must be tracked over time, for as long as the unit is deemed to continue. This goes from the moment of creation to the point where they inactive, including in the occurrences of demographic events that may cause unit to change statistical identifiers. In the SBR, a change in the statistical unit identifier does not necessarily correspond to the birth or death of the units in the longitudinal database. There are situations where a given statistical unit in the live register and snapshots may be represented by more than one distinct statistical unit over the course of its history.[8] Alternately, a newly created unit (e.g. birth) can be found at a later date, and following reception of newer information, to represent an already existing unit after new information is received (e.g. change in ownership).

Tracking of units through time might be made easier with the creation of a new *longitudinal* statistical identifier. This longitudinal identifier(s) – a renumbering – must be assigned or associated to each and every statistical unit in the SBR. The renumbering must allow for the cross sectional identification of statistical units (match or associate like units), and for longitudinal identification of surviving units (continuing, predecessor/successor). This can be referred to as a renumbering within the period and a renumbering between periods.

[7] See Geurts (2016) and Rollin (2013)

[8] The outcome of demographic events is available in the SBR guidelines and business demography manual. See also section 3.2.4 of the current guidelines.

The creation of this identifier would follow the method used to assign identifiers in the SBR. The combination of longitudinal identifier / statistical unit identifier should not be duplicated in any given reference period.

Other information useful to the interpretation of the history of the statistical units should also be included in these tables. This may include for example columns to record demographic events and business status.

Two options to store the longitudinal identifier(s) of the statistical units are presented below. Other approaches may be possible. Information can be stored into a set of reference period specific tables, or into a single longitudinal table.

Reference Period Specific Renumbering Tables

In an approach based on a series of reference period based tables, each table include current to previous and current to current reference period linked statistical units. The series of reference period based tables can be viewed as concordance tables. A table is created by reference period, and there is as many tables as there are reference periods included in the database. Statistics Netherland has adopted a similar approach for the production of business demography statistics (see Example box below).

History of units is represented by a series of linked statistical units' identifiers for each consecutive reference period (current to previous reference period). Each pair of identifier, linked units in reference period T-1 and reference T, corresponds to a continuing business, a survivor. To create a continuous image of the statistical units, predecessor/successor, requires that information from the complete series of table be put together. A statistical units that ceased existing in period T-1 will not find a link to a unit in period T, while a new statistical unit (birth) in period T will not be linked to a unit in period T-1.

Figure 3.2: Reference Period Specific Renumbering Tables

Reference Period	Longit Id *	Stat Unit ID at T	Stat Unit ID at T-1	…
T-1	L1	S1	S1	
T-1	L1	S2	S2	
T-1	L2	S3	S5	
T-1	L3	S4	S4	
…	…	…	…	

Reference Period	Longit Id *	Stat Unit ID at T	Stat Unit ID at T-1	…
T	L1	S1	S1	
T	L1	S2	S2	
T	L2	S3	S3	
T	L3	S6	S4	
…	…	…	…	

*: Note that the Longit Id column is optional.

In order to facilitate the tracking of statistical units, especially in the cases of demographic events leading to the splits or mergers of statistical units, all statistical units can be associated to a new longitudinal renumbering scheme. This new longitudinal identifier helps identifying continuing units and intra-period (current to current reference period) linkages. The L1 Longit Id in the above example tables would be cases of intra-period linkages due to a merger or a take-over.

In the example from Statistics Netherlands, a second set of inter-year table substitutes the needs for a longitudinal identifier.

The benefit of this approach is that a table can be created for each of the snapshots in the SBR, maximizing the use of the current structure of the SBR, while accepting that the snapshots are not retroactively updated with recent methodological changes and new information on statistical units.

An element to consider is the fact that statistics can only be produced for the shortest reference period covered by the unique predecessor/successor table. A shorter reference period, for example monthly, will offer more flexibility than a longer reference period (annual). Therefore, decision on the length of the reference period is critical.

Single Longitudinal Table

In an approach based on a single table, a new set of longitudinal identifiers would be associated to statistical units. Each single row of a longitudinal identifier would be associated with a reference period start and a reference period end dates (effective dates), and to a SBR statistical unit. There could be only one pair of longitudinal to SBR identifier in the table for a given period. However, the longitudinal identifier can appear more than once over the same reference period if associated to more than one

Data Model Example: Statistics Netherlands' approach to track businesses linkages

In this box a practical solution is outlined for cases for which the business register is (for example for reasons of design principles or administrative changes) not able to store linking information on enterprises that are renumbered between two frames in the business register but are in fact identical.

Renumbering can be identified for instance by the matching steps in chapter 5 of the Eurostat – OECD Manual on Business Demography Statistics.

Linking enterprises can be important in survival analysis and for assigning a correct yearly total of turnover or exports to an enterprise that during the year has two or more different enterprise identifiers (IDs). Assigning correct yearly totals is important for the analysis of growth (or decline) of individual enterprises.

The approach is to create two different linking tables. The first one is for intra-year renumbering. The second one is for inter-year renumbering.

Given this sample chain of enterprises:

2014	→	2014	→	2014	→	2015
A	→	**B**	→	**C**	→	**D**

- **This chain can be split up into two parts:**
  - 2014: **A → B → C**
  - 2014 – 2015: **C → D**
- **In order to reconstruct this chain we need two link tables:**
  - One within the same year *(for 2014)*
  - One between two different years *(transition 2014 – 2015)*

ID_Start *(2014)*	ID *(2014)*	ID_Finish *(2014)*
A	A	C
A	B	C
A	C	C

Moving forward or backward in the chain is accomplished by using the ID_start en ID_finish columns: A ultimately becomes C (in this year). Totals for all the renumbered enterprises in a year can be obtained by summation of the values for the ID column (the middle column) and grouping by either the first or last column.

The chain is elongated by the inter-year table, by coupling the ID finish columns of both tables.

SBR statistical units. The creation of the continuous image of the statistical units, predecessor/successor, is included in a single table.

Figure 3.3: Single Longitudinal Renumbering Table

Longit Id	Stat Unit	PeriodStart	PeriodEnd	...
L1	S1	2008/01/01	2012/12/31	
L1	S2	2013/01/01	9999/12/31	
L2	S3	2009/04/01	9999/12/31	
L2	S4	2016/01/01	9999/12/31	
L3	S5	2008/01/01	2015/12/31	
…				

Grouping the longitudinal identifiers into a single multi-year table reduces the number of tables in the database and offers the flexibility required for the production of business demography statistics for various reference periods (month, quarter or year).

No matter the approach chosen (single table or series of table), the table(s) of longitudinal identifiers shall include basic information on the status of the units (e.g. activity status), on the demographic event/type leading to the creation of a unit or cessation of activity, and operating and legal addresses and names.

Data Model Example – Finland

Overview

In Finland, Business Demography statistics are produced from statistical business register (SBR). A separate longitudinal database is not maintained, as the longitudinal information is integrated into the SBR. Longitudinal microdata and summary reports are produced by using a separate Business Demography (BD) table in the SBR.

The Business Demography table contains variables for:

- Enterprise ID
- Year of Birth
- Year of Death
- Indicator for enterprise birth
- Indicator for enterprise death
- Enterprise ID of predecessor
- Enterprise ID of successor (continuator)

Therefore, one row in the BD table represents one period of activity for each enterprise, and consequently one enterprise may have multiple rows in the table.

The longitudinal Business Demography data is produced annually for Eurostat and the BD table is updated for this purpose. The first two production steps update the BD table using SBR data as a source. After the updating is complete, finally a reporting data set is produced. All production steps follow the OECD – Eurostat Business Demography Manual.

The annual updating cycle of BD table corresponds to the most regular user need for longitudinal data. When a longitudinal micro data set is produced, a frozen version of the BD table is stored for the corresponding BD. Additional versions can be produced when needed.

Three steps from SBR to Business Demography

The first step in producing BD is to determine enterprise births, deaths and reactivations from annual information on enterprises in SBR. For each enterprise, years of birth and death are stored in the SBR's Business Demography table. Reactivations are also detected in this step and are updated to the BD table by simply removing the previous year of death.

The second step is to detect real enterprise births and deaths. Also, to find cases of survival by take-over, a predecessor or successor may be determined for some enterprise births or deaths. This step utilizes national administrative data and the matching methods specified in the OECD – Eurostat BD Manual. Largest enterprise births and deaths are verified and edited manually using the SBR microdata editing tool, which has a separate view for editing data in the BD table.

The results of the second step are updated into the BD table in SBR. For each enterprise birth, the birth indicator is set to either 0 for other creations, or 1 for "real" births to be included in the BD statistics. Indicator for enterprise deaths is updated similarly. In cases where a newly born enterprise can be concluded to have a predecessor, also the enterprise ID of the potential predecessor is updated. Again, for enterprise deaths, the ID of possible successor is updated similarly.

The third and final step is to create a longitudinal panel data for reporting. Business Demography table is merged with annual information from SBR to link enterprise records. Where an enterprise has been taken over by a continuator enterprise, the enterprise ID is set to be that of the continuator and year of birth is set to be that of the predecessor. The original enterprise ID is recorded as a secondary ID.

The resulting table contains annual turnover, employment, and classification variables for each enterprise. Indicators for survival and employer enterprises are calculated. This reporting table is then used as a basis for all reporting on BD.

3.2.5 Demographic Events and Characteristics

Typology of business demographic events has been covered in few reference manuals. The EUROSTAT-OECD Manual on Business Demography Statistics, the EU Business Registers Recommendations Manual and the UNECE Guidelines on Statistical Registers contain sections and chapters on this very topic.

A fully operational business demography database shall record all events having an impact on the statistical units' production factors from the date of birth to the date the unit has ceased operation. Events date must be the effective date, and if not available the closest probable date.

The SBR guidelines defines demographic events as following:

> *7.25 …an event that has an impact on the existence of a statistical unit or on links between statistical units.*
>
> *7.26 is based on changes in the existence of production factors, or in their distribution, within and among statistical units. It may involve the continuity (survival) of a unit over time or its discontinuity (death). It may also be accompanied by changes to the values of certain characteristics, such as size or type of economic activity.*

No additional recommendation on typology of events specifically related to business demography statistics is needed in the current guideline document.

The construction of the history of statistical units needs to consider all demographic events impacting statistical units. The history can be reconstructed by combining information from the live and frozen

registers and the logs[9]. All changes or updates to the statistical units in the SBR, including demographic events, should be recorded in the logs. A journal entry in the logs should list, among others, the effective date of occurrence and the type of events, and linked statistical unit(s).

The SBR may not have a record of all events impacting statistical units, for example retroactive changes may not be all recorded in the SBR. Linkages of units in consecutive frozen frames may reveal occurrences of unrecorded events. Efforts should be made to record and to categorize the newly revealed events (assign typology). Decision (birth, death continuity) should be based on comparative analysis of production factors or changes in size or classification characteristics of concerned units.

Basic decision rules should be automated for small units, making use of recommendations in the SBR guidelines and business demography manual. If the characterization of events requires manual intervention, their number should be small and focus given to the most significant statistical units.

3.3 Characteristics of the Longitudinal Database

This section details how the database is constructed, how the longitudinal information is stored in the database, the location of the longitudinal database and its relationship with the SBR, and the need to create summary tables.

3.3.1 Structure of the longitudinal database

Two broad types of structures are foreseeable for the construction of a longitudinal database from the SBR. The type of structure selected will be guided by the decision made on the longitudinal identifier (see section 3.2.4).

The longitudinal database can be a series of reference period specific data tables, or a series of data specific linkable and longitudinal data tables.

Reference Period Specific Tables

A database containing reference specific tables would work well with the option to build a series of reference period specific predecessor/successor table in section 3.2.4. Under this option, a table is created for each of the reference period covered by the database. A new table is added to the database when the reference period is completed.

Each row in a table contains all information on the longitudinal units available in the longitudinal database for the reference period. This includes unit's identification information (identifiers, names, addresses, classification variables (industry, geography…), and size variables (employment, turnover, assets…). In order to construct the history of a unit or a population, the users must link the tables for the period of interest.

Figure 3.4: Reference Period Specific Data Tables

Period	Longit Id	Stat Unit	VarA	VarB	VarC	VarD	VarE	…
T	L1							
T	L2							
…	…							

[9] Logs are records of changes or updates to information in the SBR. For example, a change to a classification variables will be supported by an entry to a log table. This entry will include the date, the source, user/process identifier, and relevant information supporting or justifying the change.

Period	Longit Id	Stat Unit	VarA	VarB	VarC	VarD	VarE	…
T+1	L1							
T+1	L2							
…	…							

This is probably the easiest design for users to query information from the longitudinal database. Creation of analytical database requires linking tables and select information of interest. However, this structure limits the shortest reference period for analysis to the reference period of the tables in the database. If tables are created annually, it is impossible to produce quarterly or monthly estimates. The number of tables grows with time and can become considerable if frequency of creation is sub-annual. Adding new information in the database, or performing retroactive update to the database implies that all or many tables need to be updated.

Information Specific Tables

A database containing information specific tables works well with the option to store longitudinal identifiers in a single table. Under this approach, the database contains issue or data specific tables. This database has as component tables as there are types of variables or data sources included in the database. The records in each table include the longitudinal identifier(s) which server as the key to link and group information from tables of interest. The content and reference period of each table is specific and relates to the data source and topic. For example, there can be a table by administrative data source, a table for each classification variables (industry, geography, activity), along with the table containing longitudinal identifiers (section 3.2.4).

Figure 3.5: Information Specific Data Tables – Size variables

Longit Id	Stat Unit	Revenue	Period
L1		$	T1
L1		$	T2
L1		$	T3
….			

Longit Id	Stat Unit	Employee	Period
L1		#	T1
L1		#	T2
L1		#	T3
….			

Figure 3.6: Information Specific Data Tables – Characteristics variables

Longit Id	Stat Unit	Industry	PeriodStart	PeriodEnd
L1		11	2008/01	2010/03
L1		22	2010/04	2015/06
L1		34	2015/07	9999/12
….				

This type of database is more complex to query than a database structured with reference period specific tables. However, it offers more flexibility. Data can be stored based on specific reference period, for example proper to each individual data source (monthly, quarterly, annual, etc.). Tables containing information that do not tend to change at every reference period, such as classification variables for example (geography, industry/activity code, etc.) can be constructed so that individual record/row cover multiple reference periods, by adding a start and end date, as shown in Figure 3.6. This

Under this structure, the number of tables in the database is relatively stable. Adding new information is simpler as calls for the creation of new tables and does not impact existing tables. Retroactive updates to existing information, including changes in methods or in classifications, is only applied to the impacted content specific tables. In order to ease data extraction, a set of pre-defined output tables could be automatically generated. Further, with this approach allows for the production of sub-annual business demography statistics.

3.3.2 Location of the SBR

There are several possible scenarios for the location of the new longitudinal database. In Section 2.10 of the SBR guidelines manual, it is indicated that "*responsibility for and control of a satellite are separated from the SBR and usually take place in a different environment.*" The argument being that "*more information can be stored and managed without interfering with the basic functions of the SBR*", and a satellite approach would "*ease the workload on SBR staff and reduce the complexity of the SBR system*". Counter arguments to support a database that has closer links to the SBR exist.

This new database must have new and dedicated resources; it cannot be supported by existing SBR resources. Because the SBR and the new longitudinal database share common data sources, are updated based on similar rules, concepts and methods, and must hold similar population counts and characteristics, there is a logic in having the new longitudinal database resides close to the SBR.

3.3.3 Relation of the database with the SBR

The inception of the LD is a significant undertaking. A new permanent table(s)/satellite database/product has to be created. This product will store statistical units' historical/longitudinal

characteristics information and size data. It will become the part of the SBR serving as the source for the production of business demography statistics.

There is no single approach to reach this objective. Paragraph 2.53 of the UNECE SBR Guidelines recognizes this challenge and includes a recommendation for the SBR to assume the role of production of economic statistics:

> *2.53 It may be advisable for statistics derived from SBR data to be based on a satellite approach. If parts, or the whole, of one or more frozen frames are extracted from the SBR and linked to data from other sources and if the resulting information is maintained outside and independent of the SBR environment, then the product is referred to as a satellite. A satellite approach (as further elaborated in Section 2.10) is a good way of addressing issues that may arise from corrections of classifications or reference periods as the base population can be well defined and coordinated…*

Indeed, a new database is needed for the production of business demography statistics. There are several options to build this infrastructure in relation to the existing SBR. NSI's should analyse all options and identify the approach that better suits their organization. The new database can be a satellite, at a distance, to the SBR, a new product with linkages to the SBR, or one could conceptualize the *transformation* of the SBR by adjusting and adding components to the liver register for it to become longitudinal. For example, the infrastructure can be:

a. Satellite Approach (UNECE SBR Guidelines), physically separated – at a distance – from the SBR.
   Using SBR rules or with its own set of rules, update process, population, etc.
b. Partially integrated to the SBR (sharing common resources or input tables)
   Longitudinal database using same/similar update processes to the SBR, sharing same input data tables for size variables, etc.
c. Fully integrated to the SBR
   The transformation of the live register into a longitudinal live register
   The live register becomes longitudinal (from a SBR to a LSBR).

Section 2.10 of the SBR guidelines lists arguments supporting the use of a satellite approach to extend functionality of the SBR. Major points raised include separated responsibilities and controls, mainly to limit overtaxing resources in charge of the maintenance of the SBR. However, technical arguments support the development of a demography statistics database that has stronger linkages to the SBR. For example, the content of the longitudinal database should match the content of the live register, and updates must replicate most of the SBR rules (date and outcome of demographic events, corrections, insertion of new data).

In order to maximize efficiency and to reduce duplications, the processes to update and maintain the SBR live from new administrative data could be modified to share information with these new longitudinal data tables. Integration of the longitudinal aspect of businesses within the live register reduces the need to develop new processes specific to the longitudinal database. As such, it is recommended to build a longitudinal database that maximise linkages to and methods with the SBR. NSI should analyse all options and identify the approach that better suits their organization.

Figure 3.7: Satellite Approach

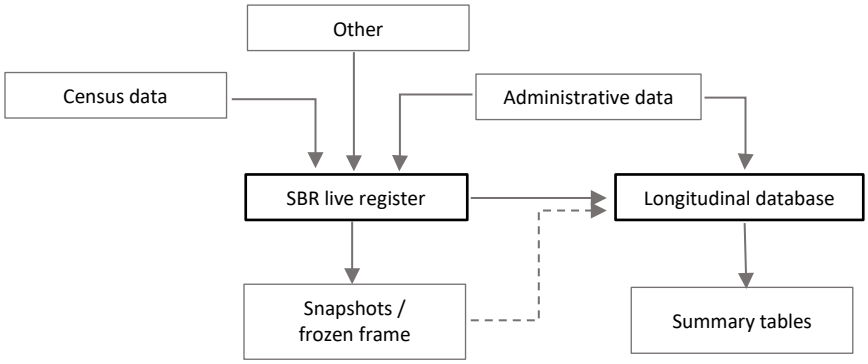


Figure 3.8: LD Partially Integrated to the SBR

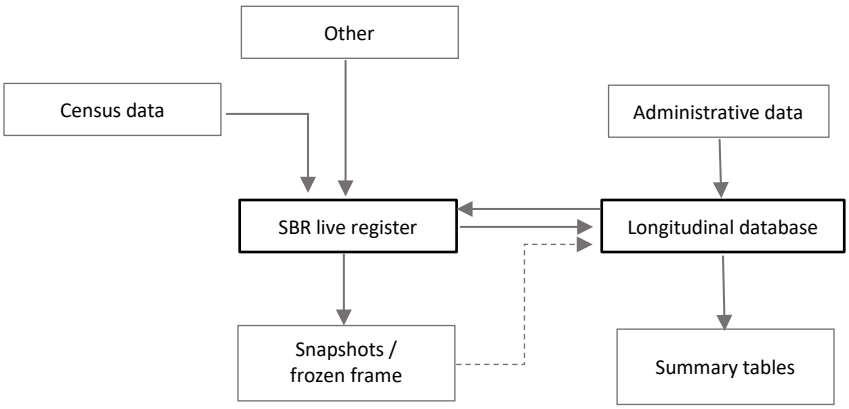
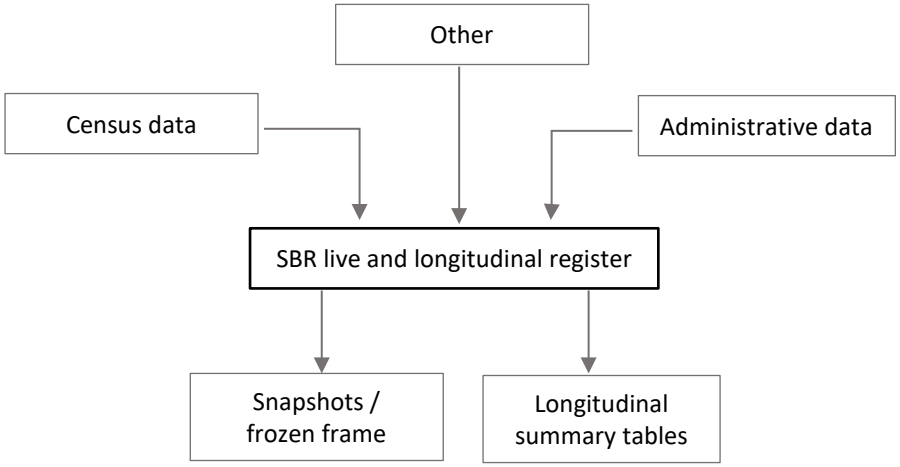


Figure 3.9: LD Fully Integrated to SBR



A LD fully integrated to the SBR (longitudinal SBR) involves significant changes to existing SBR in order to add a longitudinal component to the live register. In order to minimize burden on the SBR team, the integration of the longitudinal database with the SBR can be done in phases. For example, starting with the development of a functional longitudinal database separated from the SBR, located in the SBR environment, with a view to eventually share and merge components with the SBR. The level of integration and number of shared component (processes, input tables, etc.) can be evolving over time

as part of the regular SBR improvement and maintenance work, to ultimately get a longitudinal SBR. Ultimately, NSI should find the level of integration that fits their model best.

Not matter the approach selected, the final product must have a longitudinal aspect, align with the SBR guidelines, and match the content of the live register to the greatest possible extent. This will allow for the production of consistent and comparable demography statistics; the comparison of population at two non-consecutive points in time; and the following of individual statistical units through time.

The development of a longitudinal database with strong linkages to the SBR or the transformation of the SBR into a longitudinal SBR will allow for the production of broad range of business demography and entrepreneurship statistics by linking with other data sources.

3.3.4 Summary Tables/Snapshots

The production of business demography statistics from the longitudinal database will require to link information from multiple tables (reference period specific or information specific). This process can quickly become complex for statisticians. This is not different than with the SBR. In order to support survey programs, NSI may generate survey frames and populations from the SBR for the survey programs to use. A similar approach should be implemented for the longitudinal database. In order to help users to produce business demography statistics from the LD, a set of ready to use longitudinal tables should be created. The content, coverage, period and rate of refresh of the summary tables must be defined by users need. Users may request that copies of these summary tables be saved for a definite period.

Similar to the SBR, the longitudinal database will be continuously updated. There might be a need to keep some form of snapshots of the database on a set frequency (monthly, quarterly or annual), on top of or alternately to historical copies of summary tables. The copies (snapshots) of the longitudinal database will consume a significant amount of space. Restrictions in terms of content, years included, number of copies and duration of conservation of the snapshots/copies of the database should be applied to limit excessive space usage.

3.4 Data Sources, Coverage, Statistical Units and Population

There are many considerations in the selection of data sources, coverage, of the type of statistical units and the populations for the production of business demography statistics. Decisions vary between business demography programs and include aspects related to international comparability, level at which administrative and other SBR data sources are available, the need to produce statistics at given industrial classification and geography/regional delineation, and so on.

In order to support a wider range of business demography programs and statistics, the design of the longitudinal database should minimize restrictions on coverage, statistical units, and populations.

Whatever NSI includes in the longitudinal database, in terms of coverage, statistical units and population, the users must be fully aware of the criteria used to define the content of the database, and limitations to the production of business demography statistics by characteristics (level of statistical units, reference period, range, classifications variables, etc.).

3.4.1 Data sources

The SBR guidelines recommendations on data sources for the SBR (Chapter 6) apply in the context of the production of demography statistics. As previously indicated, basic demography statistics can be created from the SBR, given that issue of business continuity has been resolved. One of the remaining

shortcomings of the SBR is that it does not hold units' longitudinal and up-to-date characteristics and size information; it is missing longitudinal basis.

Currently, the live SBR is updated with new administrative data covering only the most recent reference period (year, month or quarter) for each different data file. Statistical units' size and other characteristics information on the SBR are constantly updated with the newly received/loaded data. Previous vintages of administrative data and information covering older reference periods may be accessible in the snapshots. However, this information is sometimes out-dated image of the past. To better support the production of business demography statistics, up-to-date longitudinal data are needed for each and every reference period of the existence of the statistical units.

As such, to better support the production of business demography, there is a need to consider business demography statistics requirements when deciding on the approach to maintain and update the longitudinal database. Size variables, mostly based on administrative data sources, and characteristics information have to be available on a longitudinal basis. For example data from administrative sources (VAT, Income Tax and Employment records) are required for a range of reference years/months for statistical units included in the longitudinal database. Longitudinal data table(s) and characteristics information have to be included in or linked to the longitudinal units' information created from the SBR.

The longitudinal data can be stored in a suite of new tables or included in a single table. For example, the content of longitudinal table can be organized based on frequency of data, or based on the data source. For example, if administrative data is available at a single frequency, e.g. annually, then a single table might work just fine for all administrative data sources. In the cases where data is reported in various frequencies, e.g. weekly, monthly, quarterly, or annual, then it might be indicated to split the longitudinal data table into separate data tables based on data source or on frequencies.

Longitudinal data tables must use the same identifier(s) for the statistical units to permit linkages. Note that administrative data sources do not offer the exact same coverage. Therefore, population covered by the longitudinal data tables will vary according to the data sources.

3.4.2 Statistical Units

One of the main focuses of business demography statistics is to produce coherent and comparable estimates of birth and death rates, or growth/decline rates by reference period, industry and geography. The feasibility of producing such statistics depends on the availability of characteristics and sizes variables for the statistical units and populations of interest. As indicated in chapter 2, a SBR supporting business demography must allow for production of statistics at the enterprise level.

NSI's developing infrastructure for the production of business demography statistics shall aim at including all types of statistical units found in the NSI's SBR (enterprises, establishments and kind-of-activity or local units), even if statistical units of lower level (establishments, KAU and LKAU) may be missing size variables information from administrative or other data sources.

Note that limitation in the availability of administrative data by statistical unit level should be documented, so that programs and data users produce business demographic statistics at the most appropriate statistical unit.

NSI could include in their development plan for the longitudinal database, provision for the allocation of administrative data from the legal entity/enterprise statistical units to the lower levels units (establishments, KAU and LKAU). This would allow for the production of business demography statistics for statistical units lower than the enterprises and at more detailed geography levels (e.g. regions and local areas).

3.4.3 Coverage and Population

Chapter 3 of the SBR guidelines state that:

> 3.2 In principle, an SBR should record all institutional units in the national economy that are engaged in productive economic activities, i.e. activities contributing to the gross domestic product…

While business demography statistics is interested in market units (above a set minimum threshold of employees or annual turnover), the longitudinal database should aim to have a similar coverage to the NSI's SBR. The longitudinal database should be as inclusive as possible and record all types of institutional units, and sectors, engaged in productive economic activities. This extended coverage provides useful information to a wider variety of programs to produce business/economic statistics from the longitudinal database. Shortcomings or limitations, in terms of quality of coverage for example, should be made available to users.

The same can be said of the population included in the longitudinal database. The SBR population include all statistical units ever created, that is the units that are alive, temporarily inactive, permanently closed, or created by error for example in case of duplications or misinterpretation of demographic events. While the longitudinal business data must account for all active businesses, based on set thresholds (turn-over and number of employees), there should be no restriction on the SBR population included in the longitudinal database. No threshold should be set for the population of units included in the longitudinal database.

The objective is for the population of the longitudinal statistical units to match as closely as possible the SBR population so to satisfy the needs of most/all business demography programs within the NSI. Similar to survey programs with the SBR, business demography programs must be able to list program specific characteristic and criteria and find a complete population of statistical units in the longitudinal database. The population selection criteria vary by and may include, for example industry classification, business size, geography and time period, among others.

3.5 LD System requirements: Update and maintenance processes, imputation, business rules and quality assurance

A longitudinal business database based on, with strong linkages to the SBR, or the addition of a longitudinal component into the SBR shall borrow as much as possible from the SBR practices in terms of updates and maintenance, correction of erroneous values in input data sources, business rules and governance, to name a few.

3.5.1 Regular/Scheduled Update and on-going Maintenance

The lack of retroactive updates to snapshots, the foregoing of new information of past reference periods, and the delay in the information received and recorded are some of the major shortcomings of the SBR to support business demography statistics. For demography statistics purposes, the update process of SBR live register must be expanded to include information for past reference periods in the longitudinal database.

The database shall be maintained, updated and improved through a comprehensive program. The longitudinal business database should be kept up to date, with the most recent information. Efforts should be made to synchronize the updates to the longitudinal database with the SBR updates, following the same rules whenever possible. Correction of errors in the SBR should also be brought into the longitudinal database.

Continuous update of the longitudinal database would minimize the number of records updated at each cycle, and may facilitate the task to categorize and record demographic events. However, the longitudinal table could also be updated on a less frequent basis, align with the reference period for which demography statistics are created. For example, an NSI producing demography statistics by quarter may only update the longitudinal database on a quarterly basis. The same applies for annual statistics.

Frequency of Updates

Once created, this new table(s) will have to be updated with new information in a similar fashion as the live register. The retroactive update of the longitudinal database is one of the major differences with the SBR update processes. With the longitudinal business database updates are no longer restricted in time. They can go back to the beginning of time or apply to any reference period in the past. The methods and rules, for example dealing with demographic events, can be the same for both databases.

Frequency of update can vary and do not have to follow the same timing as the SBR updates. The processes should ensure that for the most recent reference period the content of the longitudinal data table is similar to the information found in the live register. Data in the snapshots will differ because snapshots are frozen while the longitudinal table gets updated retroactively.

Depending on the structure and content of the database, on resources (human and IT), updates can be applied on a continuous basis up to a minimum of once a year. If possible, it is recommend updating the longitudinal database as often as possible. As such, it may be a good practice to plan and organize the updates to the longitudinal table in a similar fashion to the SBR live register update and maintenance schedule. Efforts should be made, whenever possible, to apply the same update rules to the two products (liver register and longitudinal data table).

3.5.2 Correction of administrative data and classification variables

The SBR has quality assurance rules and processes to identify and correct errors and improve quality of the SBR. The SBR processes to correct and improve the quality of its information will trickle to the longitudinal database. All changes made to the SBR by the SBR team have to be transposed to the longitudinal database. As such, there is no need to repeat the SBR rules on the longitudinal database.

The longitudinal database will be faced with its own issues. Issues that are not present in the SBR, related to the longitudinal aspect of the data. This is mostly through at the inception of the longitudinal database, when information from the SBR snapshots (not retroactively updated), are combined to the live register. The correction of issues in the longitudinal database can be split by data type: size variables (from administrative and other sources), and classification variables.

Administrative and other data sources

Treatment and processing can include correction of outlier values and imputation of missing records for example. Treatment and processing of administrative or other data sources (survey, census, etc.) may fall under the responsibility of statistical domain division(s)/group(s) in the NSI. Issues identified by the SBR domain division in the input data sources, through quality assessment rules on the SBR, should be addressed and corrected in the longitudinal business database.

Classification variables (industry, geography…)

Content, update and correction of errors in the classification variables (e.g. industry code, geography) usually fall under the responsibility of the SBR team. The team in charge of the longitudinal database

should correct issues in the classification variables that did not get corrected by the SBR, for example issues related to retroactive updates in the data.

Classification data required for the construction of the history of statistical units may not be available for all periods. There will be cases where:

- data in the earlier years might be missing;
- data gaps in the history exist;
- SBR classification logs do not cover the whole history of the units.

It might also appear that information in earlier snapshots, covering older reference periods, are based on older vintage of classifications. The longitudinal database should always include the most recent vintage of classifications. When a new vintage is deployed in the SBR, the longitudinal team shall back cast this new classification to the full history of the database. Assumptions will be required for example when an old classification code is split into two or more codes in the new classification.

Automated rules should be put in place to correct errors or to edit missing classification and values. These automated rules shall be based on agreed principles and methodologies. Whenever possible, data gaps, period with missing information, in the history of statistical units should be filled in with imputed values, if there is sufficient surrounding information to base the imputation on.

3.5.3 Dealing with changes in methods/SBR redesigns

A mature SBR may have gone through one or several redesign cycles. The redesign(s) or changes in the SBR rules, methods, or input data sources can result in changes to the structure or content of the snapshots and logs. Consequently the series of snapshots and the live register can hold tables with different formats or population, for example. Information logged in logs may also be inconsistent through time.

While it is important to give full consideration on availability and content of snapshots vintages, the decisions on the content of the longitudinal database (coverage, population, statistical units, etc.) shall mostly be guided by the structure and content of the current SBR and, if possible, any planned or foreseen changes. Given that present and future information in the SBR are usually of better quality and richer than past data, the development of the longitudinal database must have a looking forward approach, and not be limited by past information.

Creating longitudinal data with data from tables and journal of various formats brings challenges. There are few potential outcomes of SBR redesigns in the context of the longitudinal database:

Changes in SBR methodology, SBR redesign cycles, addition of new administrative data sources, or disappearing data will impact the construction of and updates to the database. The significance of the changes in the content of snapshots or logs, at a given point in time, will impose limitations in the longitudinal database. The starting point of the longitudinal database may have to match the moment of deployment of a major redesign of the SBR.

Alternately, in the case of a change in the content of the SBR due, for example, to the acquisition of a new administrative data source or in the deployment of new methods to process the input data, could translate into inconsistencies or breaks in the longitudinal availability or quality of some size variables or classification characteristics.

NSI have little control on the type of data collected by taxation offices, and changes to fiscal law, regulation and taxation requirements can lead to changes in the data included in taxation files.

Administrative data may not be available on the whole range of the longitudinal data table. Changes in the format or in the content of administrative data files may also occur from time to time.

Efforts should be made to minimize the impact of these changes or minimize breaks in the longitudinal data. Breaks in data series, changes in availability of information or improvement in methods, and others, shall be well documented. Users must know the dates or reference periods affected by the changes and the details of these changes, as changes may impact business demography statistics. Documentation of the longitudinal data table shall: indicate the vintage of the classification used, list changes in methodology and associated impacts on the data, and details on the availability of the data.

3.5.4 Business Rules and Quality

A series of business rules should be put in place. These rules are to ensure that the integrity of the longitudinal database is preserved through time, and errors are not systematically inserted during updates. NSIs shall develop rules that are specific to the content and structure of the longitudinal database.

To a minimum, the longitudinal database maintenance/update procedures should replicate the significant SBR rules. The SBR rules will ensure quality of the longitudinal database but they do not suffice, new measures are required. A series of new business rules must, for example, ensure that longitudinal units' identifiers are not duplicated, that dates do not overlap, that history of units (longitudinal identifier) contains no gap, and that the content of the longitudinal database for the most recent reference period aligns with the content of the SBR.

On top of business rules, the longitudinal business database needs to be supported by a quality assurance program. The quality rules and measures developed for the longitudinal database need not to duplicate all of the ones applied to the SBR, and should focus on aspect of the database not covered by the SBR quality assurance program.

3.5.5 Governance

Information included in the governance section of SBR guidelines (Chapter 11.3) is directly applicable to a longitudinal database based on the SBR. Sections and parts of the SBR guidelines on relationship with users, legislative framework, funding and development phase – conceptual development (pre-building phase, development phase and post-build operating phase), human resources and training can all be transposed directly to the longitudinal database.

3.5.6 IT Considerations

Again, the SBR guidelines related to IT considerations apply to the longitudinal database. The longitudinal database will share common characteristics – multiple input tables, a live copy, potentially a series of snapshots, journal tables, and so on. The IT considerations may however not be as extensive as the SBR. For example there may not be a need for a graphical user interface pointing to the longitudinal database.

3.6 Use of Longitudinal Database for Producing Business Demography Statistics

A longitudinal business database serves many purposes that cannot be filled by a SBR, by survey data, or by administrative data sources. A SBR based longitudinal business database allows for the direct production of business demography statistics and as the basis to link non-business data for the production of entrepreneurship statistics.

Numerous basic business demography statistics can be produced from the longitudinal database. Business counts, birth, death and survival rates, growth rates (employment or turn-over), by reference

period or over time, by sector/industry, region and size class are few examples of statistics that can be generated from the longitudinal database.

Further, the breadth of statistics that can be potentially derived from the SBR based longitudinal business data can be augmented through the linkage of other administrative and non-administrative data sources (owners characteristics, employee and social data) and other registers (address, population, building).

It facilitate the integration and use of administrative data sources for the production of business statistics, reducing the need to collect information through survey programs, and lowering businesses response burden.

It allows for the production of new and innovative statistics, such as the impact of social characteristics, businesses practices, government policies, environmental factors, new business level performance indicators, and much more. The NSI longitudinal business data can foster the development of new research by government and academic researchers.

3.7 Data Model example(s) (Specific countries with SBRs already developed to support DS and EIs)

NSI's have adopted several approaches to the production of business demography statistics. Some are created from the NSI's SBR, others are based on business census data, or are derived directly from administrative data sources.

This issue is discussed in section (New Zealand, Finland, Belgium, Austria, etc.).

References:

Geurts, K. (2016): *Longitudinal firm-level data: problems and solutions*, Small Business Economics, 46(3), 425-445.

Eurostat and OECD (2007): *Eurostat-OECD Manual on Business Demography Statistics*, European Commisssion, OECD Publishing, Paris, http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-010-EN.pdf.

OECD (2016), *Entrepreneurship at a Glance 2016*, OECD Publishing, Paris http://dx.doi.org/10.1787/entrepreneur_aag-2016-eng

Rollin, A.-M. "Developing a Longitudinal Structure for the National Accounts Longitudinal Microdata File (NALMF)." In *Producing Reliable Estimates from Imperfect Frames: Proceedings: Statistics Canada's International Methodological Symposium*, October 16-18, 2013. P. 306-311. Statistics Canada catalogue no. 11-522-X. Ottawa: Statistics Canada.

UNECE (2015): *Guidelines on Statistical Business Registers*, http://www.unece.org:8080/index.php?id=40574&L=0

Statistics New Zealand (2007), Improved Business Understanding via Longitudinal Database Development (IBULDD): Potential Outputs from the Longitudinal Business Database, Statistics New Zealand, Wellington.

Chapter 4: Guidance for linking the SBR with information from other sources

4.1 Introduction

Statistical business registers cover a limited set of variables on business, mostly limited to direct use relevant for the business register. Thus, key variables for a statistical business register (SBR) are economic activity - determined either as the presence of employees or turnover – and a classification of this activity. Demographic events such as mergers/acquisitions, births and deaths can also be obtained from the SBR.

Linking these data from the Business Register on the micro level (a method commonly known as Micro data linking (MDL)) to a wide variety of administrative data, other data sets, or survey data can provide new insights into the relationship between enterprise characteristics and enterprise performance. Moreover, a third dimension can be added by linking data on entrepreneurs and correlate entrepreneur characteristics with enterprise characteristics and performance.

These data can provide insights into life-cycle events of entrepreneurs and enterprises which may be relevant for policy making. These analyses can be performed using data already available in the NSI and thus do not cause any additional administrative burden on enterprises.

> In the recent years various micro data linking (MDL) exercises have been completed both in Europe as well as in other countries. MDL offers an attractive option as it can provide a cost-efficient and powerful solution to discover new information and to develop new statistics and indicators both when using existing data sets but also when combining with new data collections. Eurostat has been very active in the MDL area in the recent years; a variety of MDL projects have been performed with interesting results; some of them have already become part of a regular data production while some others start to be mature enough to follow.

4.2 Business Register Backbone

The backbone for MDL analysis is the business register, which means that all and only those units present in the business register are part of the analysis. Figure 1 shows the pivotal position of the Business Register in the coupling of data on the entrepreneur and the enterprise performance.

The size of the cylinders is a too optimistic representation of reality. Data available on either enterprises or entrepreneurs may be – at least partly – based on surveys which do not cover the entire population. Data on the level of individual enterprise which could be used to analyse economic performance (i.e. productivity) is often obtained from surveys which aim to determine macro-economic total. Therefore a stratified set-up will be used in these surveys with a focus on the larger enterprises. Especially for the smaller size-classes enterprises data may not be available.

It is recommended that the best available snapshot of the reference year in the BR is used in a micro-data linking exercise.

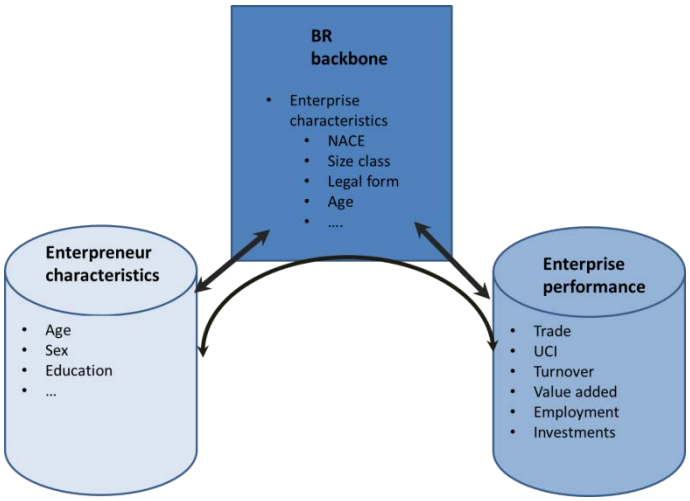


Figure 1: Central position of the Business Register in linking additional micro-data sets.

4.3 4Data available on enterprises

This section provides an overview of available data sets and options to be considered for MDL on the basis of the enterprise. This approach provides indicators on entrepreneurial performance of the enterprise.

The following paragraphs discuss in more detail the practice of linking to the most relevant available data sources on the enterprise. The section is based on the European experience, as well as specific issues noted in the Netherlands.

> Eurostat in cooperation with member states has been very active in MDL in the recent years. Several MDL projects have been developed: Trade in goods by Enterprise Characteristics (TEC), which started as an MDL project and is currently a regular production in Eurostat, Trade in services by Enterprise Characteristics (STEC), as well as projects linking Structural Business Statistics (SBS) to Trade in Goods Statistics (ITGS) as well as other Business statistics. An overview of past projects is provided in the report "Statistics on trade by enterprise characteristics – brief review of past, current and future projects linking business statistics in the European Statistical System". Current projects include the BDTEC project ("Linking Business Demography to Trade by Enterprises Characteristics") which aims at consolidating MDL between BD, trade in goods and SBS statistics. Future plans include the extension of MDL work to include Trade in Services as well.

4.3.1 Linking with Structural Business Statistics and Business Demography

In the EU, Structural Business Statistics (SBS)[i] describe the structure, conduct and performance of European enterprises. The SBS database provides basic information on the enterprise economic performance (e.g. turnover, value added, and production value) as well as information on employment, size class and other specific variables. SBS data are organised according to the NACE Rev. 2[ii] activity classification. Business Demography is a distinct dataset from SBS.

SBS data are typically based on surveys, administrative data (e.g. tax records), or a combination of both. Where SBS are based on surveys, there may not be exhaustive micro-data available for the entire population (see also the discussion in section 4.5).

Recent and ongoing Eurostat MDL projects have focussed on linking BD (and the SBR) with SBS. As regards BD the following dimensions and breakdowns were considered:

Business Demography status: Enterprise age

The adopted approach is to consider any enterprise up to and including 5 year survivals as "Young".

Survival has a specific meaning in BD, i.e. the enterprise should be active in all consecutive years (i.e. if an enterprise is a 3-year survivor in the current year then this means that it was also a 2-year survivor in the previous year, and so on). If an enterprise was not active in one year then it is not considered to have survived (however, if in the immediate following year the enterprise was "revived" then it is neither a birth, nor a death).

Very young enterprises (e.g. 0 to 1 year) are of specific interest, as these years may correspond to the so-called "born globals", i.e. enterprises that trade immediately or very shortly after their creation. Newly born and 1-year survivals can be considered in a separate sub-group.

Business Demography status: High Growth Enterprises

High Growth Enterprises where discussed in Paragraph 2.4 of the Guidelines. Specifically in the EU, Commission Regulation 439/2014 defines a collection of data on High Growth Enterprises as those enterprises "*…having at least 10 employees in t – 3, with average annualised growth in number of employees greater than 10 % per annum, over a three year period (t – 3 to t). It does not include enterprises, as defined in 11 92 0, in t – 3."*

When the Eurostat-OECD BD manual was drafted a 20% threshold had been applied for HGEs. For the purposes of analysing the trade performance of young enterprises, a more modest threshold such as 10% is probably better suited for analysis in the current economic climate (in particular, following the 2008-2009 international economic and financial crisis).

Size class: traditionally in Business Statistics enterprises are split in micro, small, medium (SMEs) and large enterprises (other groups are also possible), depending on the number of persons employed. Recently there has been increasing interest to include a breakdown between independent (enterprises that have no control links whatsoever with other enterprises) and dependent enterprises (those who belong to an enterprise group). This is relevant to size class as an "SME" that belongs to a large group will generally behave like a Large enterprise and not like a real SME.

From an economic analysis point of view, it is interesting to group separately groups of enterprises that are high growth (or not) to be able to identify factors or characteristics that may play a role in promoting (or otherwise) the occurrence of high growth. Generally speaking, it should be noted that being a High Growth enterprise is not necessarily an indication of innovativeness or a successful business model; indeed an enterprise may grow (over a specific period) due to a variety of factors: e.g. because of a successful business model, but also because of external factors outside its control, such as a general market or geographical condition that happens to favour a specific enterprise or enterprises at a given moment in time.

A specific subcategory of HGEs is the young HGEs, also known as "gazelles". Gazelles are the subset of high-growth enterprises which are up to five years old. It may be interesting to group separately gazelles and compare their performance to other enterprises, however, it should be noted that, typically, there are not so many gazelles active at a given economic activity at a given time, therefore confidentiality issues may arise.

The most commonly reported SBS characteristics in an MDL exercise are turnover, value added, employment variables (e.g. number of persons employed, number of employees and/or full-time

equivalent), wages and salaries. Other variables also can be used depending on purpose of the study. Purchases of goods and services may also be included (such as in STEC statistics for example).

> Eurostat in cooperation with the OECD have developed the entrepreneurship indicator program (EIP) that collects important indicators on enterprises and their economic performance.
>
> The TEC framework (see par. 4.3.2) provides data for the EIP indicator on the export performance of small enterprises. Another indicator, originally identified by the EIP, is the export performance of young enterprises. Data for such an indicator is not yet widely available, but there are ongoing projects aiming to produce such indicators.

4.3.2 Linking with Trade in Goods and Services data

International Trade in goods and/or services is a relevant parameter for Globalisation research and provides insight in how domestic firms are engaged in the global economy. Trade data are an interesting candidate for an MDL exercise, not only from the analytical viewpoint but also from the data source viewpoint. Several aspects of international trade can be addressed dependent on the availability of data. Interesting aspects of trade are not only value but also a breakdown in classification by product and partner country.

Trade in Goods

Specifically in the EU, Trade in Goods data with non-EU countries is comprehensively available (by product and partner country) through customs data. However, for reasons of minimizing administrative burden, the EU Member States have some limitations in the case of intra-EU trade: no breakdown of goods and partner countries is available for intra-EU trade[iii] for enterprises with trade value below a threshold, the value of which depends on the specific country.

TEC and STEC: linking trade in goods / services to the SBR

Conventional international trade statistics offer a picture of trade flows between countries, broken down by types of goods and services. While this is an important input for trade analyses, these data do not offer insights into the actors, or the types of enterprises, that are actually engaged in cross-border trade.

International trade in goods by enterprise characteristics[iv] (TEC) is a statistical domain, which unlike traditional trade statistics, aims at describing the structure of trade by characteristics of the trading enterprises, for instance by their economic activities, their size or concentration of trade.

TEC is based on linking international trade in goods (ITGS) micro data with business register (BR) information, allowing a deeper analysis of international trade by including dimensions such as economic activity, size-class of enterprises, trade concentration, geographical diversification and products traded.

TEC is an example of an MDL project which matured into a regular data production. TEC has been part of Eurostat's regular data collection since 2009. TEC provides very interesting and important statistics on the structure of trade by characteristics of the trading enterprises. As regards Entrepreneurship Indicators, a particularly interesting TEC statistic is the enterprise size class information, which can be used to provide entrepreneurship indicators on the trade performance by size class.

Eurostat together with several involved countries has also developed an experimental database for Services Trade by Enterprise Characteristics (STEC)[v]. STEC is an ongoing project by Eurostat and several countries that takes the development work done in the TEC framework forward to cover services as well. The aim of the STEC project is to produce a database on international trade in services by

enterprise characteristics (STEC). STEC data provide valuable analytical information on the characteristics of enterprises involved in the international trade of services, by breaking down traditional services trade statistics by firm size, firm ownership and firm industry. STEC data are compiled by micro-data linking the survey frame of trade in services to the national business register. STEC is

Together, STEC and TEC will provide key insights on international traders engaged of goods and services.

STEC may be a quite important data source for further MDL work. One hypothesis is that young enterprises are more likely to engage in trade in services than trade in goods (the assumption being that - except for resellers of goods - trade in goods necessitates a production process which may be difficult for a young enterprise). Although STEC itself is an MDL project, trade in services data have not been included in other MDL projects (for example, there are not considered in BDTEC). However, inclusion of trade in services data in future MDL projects is noted as an interesting next step.

Traders vs. non-traders

From the point of view of analysing the drivers of entrepreneurship, one goal of any MDL project linking to trade is to be able to split the entire population into enterprises that take part in international trade (of goods and/or services) — being an exporter, importer or two-way trader — and non-traders: enterprises who are active on domestic markets only.

Exporters (including two-way traders) are of special interest for policy makers because of their potential job creation due to demand from markets abroad. Importers (again including two way traders) are also important since they access themselves (or make available for others) raw materials, intermediate goods, services and technologies or final products that may not be available in a national market (or they may be available at better quality / lower price in an international market).

However, care should be exercised when deciding if an enterprise is a trader or not: if all enterprises that happened to sell (or buy) some goods – regardless of the value - on an international market are included in the "traders" group, then traders will also contain enterprises which are not trading because of their business model but just happened to sell (buy) something by chance, or for specific, one-off reasons (e.g. an one-off sale of used machinery abroad). Allowing such "occasional" traders to be counted as traders may cause difficulties in interpreting the results.

Therefore, in order to be able to focus only on enterprises that trade because of their business model, some assumptions must be made and some thresholds applied.

To our knowledge, there has been no in-depth international study on how to choose specific thresholds to decide whether an enterprise should be considered as trader or not. It is recommended that thresholds already used in previous projects can be adopted, to maintain comparability. Exporters, importers and two-way traders can be defined as follows:

- ***Exporters***: an enterprise that has an export value ≥ 5000 euro in the reference year and export intensity (exports / turnover) ≥ 5% is considered to be an exporter for that reference year;
- ***Importer***: an enterprise that has an import value ≥ 5000 euro in the reference year and import intensity (imports / purchases of goods and services) ≥ 5% is an importer for that reference year;
- ***Two-way trader***: an enterprise that, following the above definitions, is both exporter and importer.

The above thresholds have been applied in a number of projects by Eurostat and European countries and have been shown to provide reasonable results.

An associated difficulty is that in order to be able to apply such thresholds the turnover and purchases of goods and services values must be known and available for all enterprises. This may not always be possible (see also the discussion on handling sampling limitations).

In general, further study may be needed to explore and quantify the effect of the chosen thresholds. The country situation and economic field / market may also be taken into account.

In any case, choosing some thresholds to decide if an enterprise is a trader or not should be preferable to accepting that all enterprise who happen to buy (sell) are automatically considered as traders, for the analytical purposes of Entrepreneurship indicators.

The above approach means that the group of "non-traders" will also exhibit some (relatively low) amounts of export (import) activity in the reference year. This is the sum of the activities of enterprises whose export (import) value was lower than the thresholds. If there are trade values which cannot be linked to a known enterprise then an "Unknown" group can also be created. Therefore, comparability with total export (import) values in international trade databases can be ensured as well (i.e. the total trade may be given as the sum of trade by "traders", "non-traders" and enterprises of unknown status). It is recommended that the concepts of trader vs. occasional trader and non-trader as defined above are explained to the users, for example, when disseminating results on indicators.

Fiscal Units – Statistical Units

Data collection on trade is often performed on administrative data and survey data concerning fiscal units. For further use in data-linking analysis these data have to be linked and subsequently aggregated to the level of the statistical unit of the Business Register. Only after these steps trade characteristics can accurately and correctly linked to other enterprise characteristics. In practice coupling the data from the tax-authorities will in many cases be straightforward with a one-to-one or n to 1 linkage. More complex linkage cases and problems are described in chapter 5 of the Eurostat Manual on TEC.[vi]

Data Consistency of Large Enterprise Groups

In Statistics Netherlands all available data on enterprises belonging to the largest and most complex Enterprise Groups are analysed in conjunction. Inconsistencies between different sources and surveys are, often in consultation with the enterprise, eliminated. The goal of this approach is to ensure consistency of output at the macro-level, by intervention at an early stage in the statistical process, but is also a great advantage for research using micro-data analysis and micro-data-linking.

The following issues are described:

- Intra-annual business demography changes

  Intra-annual demography events are challenging as monthly available trade data has to assigned to an enterprise with characterises valid for a full year.
- Large and complex businesses

  The linkage may not always provide an expected outcome; trade flows may be allocated to enterprises whose characteristics seem to be in contrast with the economic reality or be incoherent with other statistics. These problems are more likely to exist for large and complex businesses.
- Incomplete business register data;

  This topic deals with the situation where a link can be established with the Business Register but its information is incomplete, for instance lacking NACE-code or size class.
- Treatment of estimated trade data

  Estimated data refers to non-response in trade data. If an estimated value can be coupled to a trader and subsequently to an enterprise this value should be used.
- Non-established traders

  Non-established traders are foreign companies which carry out trade transactions in the

reporting Member State and are registered for VAT or have appointed a tax representative. Entities without permanent establishment are not included in the Business Register.

- VAT-groups

  In the EU, under certain circumstances, several VAT declarants may report VAT as a group. In this
  case, VAT is recorded only by one VAT number. The contents of a VAT group may correspond to more than one enterprise. In these cases where there is 1-to-*n* (or even a *n*-to-*n*) relationship between VAT-identification and enterprises, the allocation of trade over different enterprises has to approximated. Suggested approaches include

  1. allocation of all trade value to the dominant enterprise
  2. allocation of specific products codes to the enterprises with the activity code that closely matches the NACE-code of the enterprise

4.3.3 Group status, size class and Foreign ownership

Availability of the enterprise size-class information and group status is essential for the entrepreneurship indicators as it allows comparisons between independent enterprises and dependent enterprises (i.e. those that belong to a group), Micro, Small and/or Medium Enterprises to Large ones, etc. Currently in European business and trade statistics size-class information is based on the number of persons employed by the enterprise. There is no consideration on whether the enterprise belongs (or not) to an enterprise group.

Group status is also very important for the analysis, as it is well known that dependent enterprises may behave in a different way (especially those who are part of a multinational group). Furthermore, in the EU, for policy purposes the definition of an SME considers the information on the whole enterprise group (and not only the enterprise).

Foreign ownership is another interesting characteristic as foreign controlled enterprises may exhibit specific behaviour (e.g. tend to employ higher rates of personnel, yield higher value added, etc.). Foreign ownership of an enterprise group is determined by the country of residency of the ultimate controlling institutional unit as outlined in the Eurostat FATS Recommendations Manual[vii]. This characteristic is subsequently assigned to all underlying enterprises of the enterprise group. In the absence of information indicating foreign control of a group, , it may be assumed that the group is domestically controlled. The results of such an approach will obviously depend on the robustness and completeness of the data sources used for the determination of the UCIs.

4.3.4 Linking with R&D and Innovation data

Expenses on Investments or Research and Development or Innovation may influence enterprise performance. To be able to compete in a dynamic international setting it is vital for enterprises to develop new products and/or more efficient means of production.

By linking available data-sets a comprehensive analysis can be performed on the effect of international orientation (from international trade data and foreign ownership data) and Innovation/R&D/investments on enterprise performance (for example from SBS).

This data is collected on the level of the enterprise which allows for straightforward coupling via the business register. Of course, not all units will be part of the survey population. In section 4.5 problems and solutions for this issue will be discussed.

4.4 Data available on entrepreneurs

This section focus on survey data and administrative data that is available on entrepreneurs. It first defines an entrepreneur and then describes the populations from which data is collected.

4.4.1 Definition of entrepreneur

The first of goal of this section is to define the population of entrepreneurs. An approach is to define entrepreneurs as a subset of persons who hold a self-employment job according to the ICSE-93 definition[viii].

> "Self-employment jobs are those jobs where the remuneration is directly dependent upon the profits (or the potential for profits) derived from the goods and services produced (where own consumption is considered to be part of profits). The incumbents make the operational decisions affecting the enterprise, or delegate such decisions while retaining responsibility for the welfare of the enterprise. (In this context "enterprise" includes one-person operations.)"

Translating this definition to a business register approach would indicate that all persons who own and control an enterprise with the intention of generating income out of it are considered entrepreneurs. For different legal forms the 'own and control' criterion is discussed below:

- One-person operations/sole proprietorship of an enterprise with no legal form: this form has all characteristics defined in the ICSE-definition.
- Partnership. In this case there are at least two entrepreneurs, with combined responsibility for the welfare of the enterprise and all partners have a remuneration which is dependent of the profits generated. Each of the partners is considered entrepreneur. Therefore, if no choice is made which is the most relevant entrepreneur the number of entrepreneurs will be higher than the number of enterprises.
- Limited liability Corporation. The owner-manager of a limited liability corporation is defined as a special case in the ILO-classification.

> *Owner-managers of incorporated enterprises are workers who hold a job in an incorporated enterprise, in which they: (a) alone, or together with other members of their families or one or a few partners, hold controlling ownership of the enterprise; and (b) have the authority to act on its behalf as regards contracts with other organizations and the hiring and dismissal of persons in paid employment with the same organization […].*

In the system of national accounts[ix] the group of self-employed persons is limited to those who own an enterprise without legal form and the remuneration of owner-managers of an incorporated enterprise is considered as wage. Therefore this group is classified as employee.

Owner manager of limited liability companies in the Netherlands: Employee or Entrepreneur?

Statistics Netherlands considers owner-managers of limited liability companies as entrepreneurs and not as employees even though there may by a labour contract and part of their income from the limited liability company is taxed as wages.

For taxation purposes a fictitious wage has to be paid by the limited liability company to the owner-manager. This fictitious wage (at least €45000) is taxed as wage and avoids tax arbitrage with respect to choice of the legal form. In general it is more profitable to transfer the profits from the enterprise to the owner by means of dividend instead of via wages. This is not possible for non-limited companies and thus would favour owners of enterprises with specific legal form .

With respect to taxation owners-managers of limited liability companies are defined as persons who have a decisive vote (with or without their partner or spouse) in the general meeting of shareholders with regard to the continuation of their labour agreement with the limited.

For this reason owner-managers are not entitled to unemployed benefits and do not pay social charges for unemployment.

In conclusion, owner-managers are for taxation reasons treated as employees, but are in the social security system not considered as employees. Moreover, in daily practice owner-managers are as much considered to be an entrepreneur as are owner of non-limited company. Also, self-assessment in the labour force survey shows that they classify themselves as entrepreneurs, not as employees.

One could argue, however, that the group of persons that is dependent of the income generated by an enterprise with legal status acts in in a similar way. They control the enterprise. In a strict juridical meaning they do not own the limited as they only own the shares of the limited. However, the income generated from the enterprise is in a similar way dependant on the welfare of the enterprise. In conclusion, owner-managers of a limited company are not different from owners of a non-limited company: they manage and (indirectly) own the enterprise, and their remuneration is dependent of the welfare of the enterprise.

Therefore it is best to attribute this group to self-employed persons/owner-managers and thus as entrepreneurs, but also consider it as a separate sub-group so that users of statistical information can select a specific group dependent of research or policy interests.

4.4.2 Data on Entrepreneurs

The first entry-point for data on entrepreneurs is the Trade Register which is maintained by the Chambers of Commerce. One of the goals of the Trade Register is to create transparency in the economic traffic, and thus to identify enterprises and its owners. Via coupling by means of an identifying key on data collected in a Civil Register, maintained by the government, a wide range of information with respect to the entrepreneur will be available. Furthermore, using data from the tax authorities additional financial information becomes available. Data available from the labour force survey provide insights on several aspects of the position of an individual in the labour force. This is a sample based survey so it only provides answers for a part of the population and it is time-dependant. The situation may not be relevant for the time-period the research focusses on. Finally, administrative data from the social security database will gain insight in the – if any- employer history of the entrepreneur.

Table 1 summarizes the data-sources and characteristics provided by it.

4.4.3 Populations of entrepreneurs

For different reasons the populations of the data-sources do not entirely overlap. This section focusses on the backgrounds which cause that the population of entrepreneurs from the business register is larger than the population in the 4 datasets from table 1.

Reasons why de BR population is smaller

- The limiting factor in the entrepreneur population based on the BR is the criterion that the enterprise is recorded in the Business Register. As the BR is the backbone of all statistics described in this manual an entry in the BR is a prerequisite for being considered an entrepreneur. From this limitation differences will occur compared to figures from socio-economic publications on the number of entrepreneurs.
    - Socio-economic figures on entrepreneurs are based on a survey with self-assessment with respect to the qualification of self-employment. Especially in cases where there is no obligation for an entrepreneur to enlist in the Trade Register the number of entrepreneurs in the BR will be lower.
    - No economic activity could be observed and therefore the enterprise is not included in the business register.
        - No employees
        - No VAT or turnover data is available because of exemption to provide data or the VAT value is below a certain country specific threshold
- An entrepreneur does not have to be an inhabitant of the country the enterprise is resided. In the case an inhabitant resides in a country but its enterprise is situated in another country this entrepreneur is not considered in the business register.

Reasons why the BR population is larger

- The classification in labour force survey is based on the main source of income, employment *vs* entrepreneurship. This is no criterion in the business register based approach as all persons who are owner-manager are considered an entrepreneur.
- In the labour force survey high and low age cut-off are applied in the definition of persons who are regarded as to belong to the labour force population. In the business register approach there are no such limitations.
- The labour force population of a specific country includes only those persons who are resident of that country.

Table 1 shows examples of sources of information on entrepreneurs. An entrepreneur may be part of 1 or more of the populations given in table 1 but is not necessarily part of all 4 of them.

Subject/Population	Source	Characteristic
Inhabitant	Civil Register	Address
		Age
		Sex
		Marital status
		Parents
		Children

		Nationality
		Place/ country of birth
Tax-payer	Income Tax Register	Height of income
		Source of income (employment, entrepreneurship, capital)
Labour force population	Labour force survey	Educational level
		Time spent as entrepreneur
Employee	Social Security Administration	Employee status
		Pre-entrepreneur Nace-activity

Table 1: Examples of characteristics available on entrepreneurs via linking of several administrative sources or surveys.

4.5 Data from surveys – Limitations and solutions for MDL

When data obtained from the sources mentioned above are coupled to the Business Register and the BR population is considered the reference backbone, then not for all units present in the BR data will be available. This may especially be the case for smaller enterprises in surveys as the focus is of many surveys is to obtain a complete picture of the largest most dominant enterprises, and therefore low inclusion weights for the smaller enterprises. An overview of available methods to deal with this problem is given by De Waal[x] of which two of them are described below.

Cross tabulation of characteristics derived from different sources is relying on overlap of units in both sources. Missing data can usually not be ignored because they are selective, and because of low overlap - especially in the lower size-classes - lead to large uncertainties.

Repeated weighting.

Repeated weighting (RW) is a technique that was developed at Statistics Netherlands to solve the problem of inconsistencies among tables of estimates based on multiple data sources[xi]. A separate set of weights is assigned to individual sample units for each cross table for which population totals have to be estimated. The tables that need to be estimated are estimated sequentially using in each block as many sample units as possible, in order to keep the sample variance as low as possible. The combined data from administrative data sources and surveys are divided into rectangular blocks. Such a block consists of a maximal set of variables for which data on the same units has been collected.

The availability of data determines how a table is estimated. Data from an administrative data source which covers the entire population can simply be counted. Data from surveys are weighted by means of regression weighting. In that case weights must be assigned to all units in the block to be weighted. For a survey one usually starts with the inverse inclusion probabilities of the sample units. These weights are then further adjusted by calibrating them to previously estimated totals. For a data block containing the overlap of two surveys, one usually begins with the product of the standard survey weights from each of the surveys as starting weight for each observed unit, and then corrects these starting weights by calibrating to totals known from administrative data sources and previously estimated totals.

Advantages and disadvantages of Repeated weighting

+ micro-data consistency is maintained (e.g. edit relations between variables of the same unit)

- generally unsuitable for estimation for small sub-populations

Mass-imputation

The mass imputation approach imputes values for all variables for all units in the population where no value was observed[xii]. In this case the population is the entire business register population. This leads to a rectangular data set in which the entire population is represented and for all variables a value is obtained. Consistency with published results on the same population is ensured by scaling to this population total. After mass-imputation for each individual enterprise a value is calculated for the relevant variables. For this approach to be successful it is necessary to capture all relevant variables and the relationships between them. When the relationships between variables also hold for sub-groups of the population it is straightforward to compute totals for these sub-groups summing the values for the specific variables of these sub-groups.

Advantages and disadvantages of Mass Imputation

+ some multivariate imputation methods can make better use of the information in the linked data

+ very flexible method, as different imputation methods may be used for different variables

+ suitable for estimation for small sub-populations, especially when appropriate imputation models are used

-a lot of work can go into the specification and checking of imputation models

- mass-imputed datasets can easily be abused by analysing relationships that were not sufficiently accounted for in the imputation models.

4.6 Conclusions and future work

The Business Register plays a central role in linking results from several surveys and administrative data. Data sources with a different unit of observation (for instance international trade) have to be transformed to the statistical unit of the business register. These datasets linked on the micro-level can address a wide variety of research questions and policy issues. Combining data from enterprises with data from entrepreneurs relates performance of enterprises (for example measured in terms of profits, turnover, investments, and trade volume) with characteristics of the entrepreneur.

It is not only possible to perform static analysis. Data on the history of an entrepreneur is available in these linked datasets. Answers to policy relevant issues regarding factors which discriminate between successful enterprises and entrepreneurs can be found using these micro-data linked approaches. The business register is at a key position in the centre of a data sources and allows for optimal use of already existing data.

Future developments

It would be interesting to link SBR and BD data also with international trade in services data (ITSS) or trade in services by enterprise characteristics data (STEC). STEC is an ongoing project by Eurostat and several countries that takes the development work done in the TEC framework forward to cover services as well. The purpose of the STEC project is to produce a database on international trade in services by enterprise characteristics (STEC). STEC data may also be used to obtain new Entrepreneurship indicators for enterprises engaged in international trade in services (ITSS) as well (currently available indicators take into account traders in goods only). Results from the earlier MDL projects show that most of the exporting enterprises are also importers and vice versa. In addition, the results suggest that most of the trading enterprises are involved both in goods and services trade. STEC is also particularly interesting as it is expected that many young enterprises first engage in trade in services and, some of them may move later on to trade in goods as well. Therefore, it would also be interesting to produce indicators on those enterprises that are involved in both trade in services and goods vs. those who are involved in goods only (or services only).

Another interesting topic is linking Business Demography and SBS variables to international sourcing surveys. The first goal would be to identify internationally sourcing enterprises in SBS and to analyse whether the international sourcing (i.e. movement of existing jobs) has, as expected, reduced the employment in these enterprises. Furthermore, one would analyse the effects of international sourcing on profitability and value added creation. An additional, more complicated, goal is to identify those enterprises who are exporting (exclusively or mainly) towards their parent enterprise (intra-group trade). This may be a result of international sourcing (e.g. multinationals organising production in specific countries only) or it may occur as a result of Foreign Direct Investment.

Chapter 5: Examples of non SBR approaches, for generating business demography statistics and entrepreneurship statistics

5.1 Introduction

Along the previous chapters, the importance of generating business demography and business entrepreneurship statistics has been established, describing the infrastructure and procedures needed to generate these kinds of statistics using the SBR. Nevertheless, there are countries that do not generate these statistics yet, nor have a SBR (or their SBR is not yet consolidated).

With the purpose of offering some alternatives to those countries that do not generate these statistics and that don't have a consolidated SBR, this chapter describes a few examples of some countries, based on other data sources: censuses, surveys and administrative registers, or combinations of them.

These examples can be assessed by other countries, and may consider the possibility of adapting the described methods for generating these kinds of statistics by using their own resources and available infrastructure.

The examples presented in this chapter correspond to Mexico, South Korea and the United States, all of them describe the type of statistics generated, and both the data sources and methodology used.

The objective in Mexico's example is to generate business demography statistics by using data collected in Economic Business Censuses, which are held every 5 years. In this example, the statistical unit is the establishment.

South Korea produces business demography statistics by linking various sources of administrative data; such statistics are related to births, deaths, and survivals of businesses; they also encompass high growth and gazelle enterprises. In this example, the enterprise is the statistical unit.

The United States presents an example in which data collected by an annual survey is combined with administrative records to produce enterprise and establishment level statistics, which describe the characteristics of businesses and business owners.

The following pages describe these examples with more detail, explaining the methodology employed and the results obtained.

5.2 Countries Examples

5.2.1 Mexico's example: Demography of businesses based on the Economic Censuses in Mexico

I. Mexico´s example: Demography of Businesses based on the Economic Censuses in Mexico

Mexico has held Economic Censuses each five years since 1930; this has helped to construct an important statistical infrastructure, from which Business demographic statistics in Mexico have been generated. The observation unit in these studies is the establishment.

This section gives a general overview of a study conducted in Mexico to generate Business demographic statistics, describing general characteristics of the study, methodology employed, obtained results, and present and future work.

1 Purpose of the study

Obtain the main indicators of survival, mortality and life expectancy that express the demographic dynamics of the business population in Mexico, through the use of applied demographic techniques on data captured by the Economic Censuses.

2 General characteristics of the study

There is a Statistical Business Register in México, its objective is to be the backbone of all the economic statistics production processes, as well as the basis for longitudinal and demographic studies, nevertheless since the Statistical Business Register is still in its initial stage, the study on business demography was based on the results of the Economic Censuses that are held every five years in Mexico, which capture all the businesses of all sectors of the country's economy, except for the agricultural sector.

The five-year periodicity of the Economic Censuses has allowed the construction of a large statistical infrastructure, since there is economic data for each of the businesses that have existed and exist throughout the country. For this study we used the results of six Economic Censuses, carried out from 1989 to 2014. In each of them we have the year of birth of every business, therefore, this variable was the basis for the study.

The year of birth of every business allows knowing how many years it has survived every time a new census is held. For example, if a business was born in 1983, and it was observed in the 1989 census, it would have survived 6 years. If it was still alive in the 1994 census, it would have survived 11 years, and so on until observing that if it was still alive in the 2014 census it would have survived for 31 years.

Thus, with the data on the year of birth, and considering all the businesses born in the same year, that is, all businesses of the same generation, we obtain data on how many survivors from that generation there are every five years (every census) and therefore how many deaths. With this data the probabilities of survival and death of every generation at each age are calculated. Now, considering all the generations, data on the average of survival and death can be obtained according to the age of the businesses and, hence, the chances of surviving at birth, of surviving or dying since they are already x years of age, life expectancy in future years at birth, and life expectancy in future years at any age.

As already mentioned, the data on six censuses was used for the study: 1989, 1994, 1999, 2004, 2009 and 2014; 26 generations were considered, from the ones born in 1983 to the ones born in 2008. This presented a volume of about 20 million analyzed records.

The results of the study are Survival and Mortality Tables on businesses; these tables contain the following for every age *x*: A) the number of surviving businesses *S(x)* at the end of age *x*, B) probabilities *p(x)* of businesses living *x+1* years as of their birth, C) probabilities *q(x)* of businesses dying before reaching age *x+1*, and D) business´ future life expectancy in years at age *x*, *E(x)*, that is, the number of years to live as of age *x*.
These tables are presented with different levels of disaggregation: geographic (National or by state); by economic sector (Manufacturing, trade, private and non-financial services and total of sectors) and by the size of the establishment (stratum of employed personnel).

3 Methodology

In the study, the first generation (*y = 1983*) was formed with the businesses that were 5 years old in the 1989 census, S_{1989} (5), the second generation (*y=1984*) with the businesses that were 4 years old in that moment, S_{1989} (4), and so on. In the next census edition 1994, births that occurred during the inter-census period were captured, that is, businesses that were 0,1,2,3 and 4 years old in that moment, so monitoring of those generations started as of their date of birth, quantifying the surviving establishments in every census, until reaching the 2014 census. Work continued in the same way with the following census data, counting the surviving establishments of each generation and forming new annual generations with the births detected, as shown in the following table:

Genera-tion (y)	Economic Censuses											
	1989		1994		1999		2004		2009		2014	
	Age (x)	Number of Businesses S(x)	Age (x)	Number of Businesses S(x)	Age (x)	Number of Businesses S(x)	Age (x)	Number of Businesses S(x)	Age (x)	Number of Businesses S(x)	Age (x)	Number of Businesses S(x)
1983	5	$S_{1989}(5)$	10	$S_{1994}(10)$	15	$S_{1999}(15)$	20	$S_{2004}(20)$	25	$S_{2009}(25)$	30	$S_{2014}(30)$
1984	4	$S_{1989}(4)$	9	$S_{1994}(9)$	14	$S_{1999}(14)$	19	$S_{2004}(19)$	24	$S_{2009}(24)$	29	$S_{2014}(29)$
1985	3	$S_{1989}(3)$	8	$S_{1994}(8)$	13	$S_{1999}(13)$	18	$S_{2004}(18)$	23	$S_{2009}(23)$	28	$S_{2014}(28)$
1986	2	$S_{1989}(2)$	7	$S_{1994}(7)$	12	$S_{1999}(12)$	17	$S_{2004}(17)$	22	$S_{2009}(22)$	27	$S_{2014}(27)$
1987	1	$S_{1989}(1)$	6	$S_{1994}(6)$	11	$S_{1999}(11)$	16	$S_{2004}(16)$	21	$S_{2009}(21)$	26	$S_{2014}(26)$
1988	0	$S_{1989}(0)$	5	$S_{1994}(5)$	10	$S_{1999}(10)$	15	$S_{2004}(15)$	20	$S_{2009}(20)$	25	$S_{2014}(25)$
1989			4	$S_{1994}(4)$	9	$S_{1999}(9)$	14	$S_{2004}(14)$	19	$S_{2009}(19)$	24	$S_{2014}(24)$
1990			3	$S_{1994}(3)$	8	$S_{1999}(8)$	13	$S_{2004}(13)$	18	$S_{2009}(18)$	23	$S_{2014}(23)$
1991			2	$S_{1994}(2)$	7	$S_{1999}(7)$	12	$S_{2004}(12)$	17	$S_{2009}(17)$	22	$S_{2014}(22)$
1992			1	$S_{1994}(1)$	6	$S_{1999}(6)$	11	$S_{2004}(11)$	16	$S_{2009}(16)$	21	$S_{2014}(21)$
1993			0	$S_{1994}(0)$	5	$S_{1999}(5)$	10	$S_{2004}(10)$	15	$S_{2009}(15)$	20	$S_{2014}(20)$
1994					4	$S_{1999}(4)$	9	$S_{2004}(9)$	14	$S_{2009}(14)$	19	$S_{2014}(19)$
…					…	…	…	…	…	…	…	…
2008									0	$S_{2009}(0)$	5	$S_{2014}(5)$
2009											4	$S_{2014}(4)$
2010											3	$S_{2014}(3)$
2011											2	$S_{2014}(2)$
2012											1	$S_{2014}(1)$

2013											0	$S_{2014}(0)$

Once the previous process is completed for each of the 26 generations that have observations for at least two census moments (from generation 1983 to generation 2008), the five-year probabilities of survival were calculated with this table, as explained in the following section.

Calculation of survival and death probabilities, and life expectancy of businesses

The proportion of survivors from one census to another of each generation were calculated, dividing the number of survivors of a census by the number of survivors of the immediate previous census, as shown in the following table:

Generation	1994/1989		1999/1994		2004/1999		2009/2004		2014/2009	
	Age	Prop	Age	Prop	Age	Prop	Age	Prop	Age	Prop
1983	5	$S_{1994}(10) / S_{1989}(5)$	10	$S_{1999}(15) / S_{1994}(10)$	15	$S_{2004}(20) / S_{1999}(15)$	20	$S_{2009}(25) / S_{2004}(20)$	25	$S_{2014}(30) / S_{2009}(25)$
1984	4	$S_{1994}(9) / S_{1989}(4)$	9	$S_{1999}(14) / S_{1994}(9)$	14	$S_{2004}(19) / S_{1999}(14)$	19	$S_{2009}(24) / S_{2004}(19)$	24	$S_{2014}(29) / S_{2009}(24)$
1985	3	$S_{1994}(8) / S_{1989}(3)$	8	$S_{1999}(13) / S_{1994}(8)$	13	$S_{2004}(18) / S_{1999}(13)$	18	$S_{2009}(23) / S_{2004}(18)$	23	$S_{2014}(28) / S_{2009}(23)$
1982	2	$S_{1994}(7) / S_{1989}(2)$	7	$S_{1999}(12) / S_{1994}(7)$	12	$S_{2004}(17) / S_{1999}(12)$	17	$S_{2009}(22) / S_{2004}(17)$	22	$S_{2014}(27) / S_{2009}(22)$
1987	1	$S_{1994}(6) / S_{1989}(1)$	6	$S_{1999}(11) / S_{1994}(6)$	11	$S_{2004}(16) / S_{1999}(11)$	16	$S_{2009}(21) / S_{2004}(16)$	21	$S_{2014}(26) / S_{2009}(21)$
1988	0	$S_{1994}(5) / S_{1989}(0)$	5	$S_{1999}(10) / S_{1994}(5)$	10	$S_{2004}(15) / S_{1999}(10)$	15	$S_{2009}(20) / S_{2004}(15)$	20	$S_{2014}(25) / S_{2009}(20)$
1989			4	$S_{1999}(9) / S_{1994}(4)$	9	$S_{2004}(14) / S_{1999}(9)$	14	$S_{2009}(19) / S_{2004}(14)$	19	$S_{2014}(24) / S_{2009}(19)$
1990			3	$S_{1999}(8) / S_{1994}(3)$	8	$S_{2004}(13) / S_{1999}(8)$	13	$S_{2009}(18) / S_{2004}(13)$	18	$S_{2014}(23) / S_{2009}(18)$
…			…	…	…	…	….	…	…	…
2008									0	$S_{2014}(5) / S_{2009}(0)$

These proportions (Prop) come close to the probability of reaching age x + 5, since in the previous census the business was age x, reason why they are an estimate of the probabilities of quinquennial survival.

Using these five-year survival probabilities, the estimation of annual survival probabilities was done for every generation, through an interpolation procedure, obtaining annual rates based on known points (quinquennial). With these estimates of annual survival probability, a table was constructed considering an initial population of 100,000 businesses for every generation, with the following structure:

Age	26 generations				
	S(x) 1983	S(x) 1984	…	S(x) 2007	S(x) 2008
0	S(0) 1983 =100,000	S(0) 1984 =100,000		S(0) 2007 =100,000	S(0) 2008 =100,000
1	S (1) 1983	S (1) 1984		S (1) 2007	S (1) 2008
2	S (2) 1983	S (2) 1984		S (2) 2007	S (2) 2008
…	…	…	…	…	…
30	S (30) 1983	S (30) 1984	…	S (30) 2007	S (30) 2088

This table contains 26 annual survival vectors (columns) of annual survival (one vector per generation). The vectors were calculated up to the age of 30 years, which is the limit of years for the first generation (1983) to the most recent observation (2014).

In these annual tables of survivors each year, out of the 100,000 initial businesses of every generation, it was observed that as the age *x* increases, the number of surviving businesses is decreasing but tends to stabilize, that is, there comes a time for each generation where practically all businesses survive.

The next step was to convert the previous table (which contains specific survivor numbers for each of the 26 generations) to a single table that reflects the average of survivors for every age, but regardless of the generation, to reach a table that is similar to the ones used in studies on human demography.

For the above, using the data from the annual survival tables, the average of survivors of the 26 generations was calculated for each age *x*, resulting in a table with the following structure:

Age	**S(x)**
0	$\sum_{y=1983}^{2008} \frac{S(0)\ \ y}{26}$
1	$\sum_{y=1983}^{2008} \frac{S(1)\ \ y}{26}$
…	…
30	$\sum_{y=1983}^{2008} \frac{S(30)\ \ y}{26}$

From the results obtained in this table, the survival behavior was modeled through a linear adjustment, with very favorable results.

4 Obtained results

Using the survival function, Survival and Mortality Tables were obtained with the following information:

Age x	S(x)	p(x)	q(x)	E(x)

where:

x: Age of businesses

S(x): Number of surviving businesses at the end of age (x).

p(x): Probability of a business surviving during age (x).

q(x): Probability of a business dying before reaching age (x).

E(x): Life expectancy at age x.

furthermore:

$$p(x) = \frac{S(x+1)}{S(0)},$$

$$q(x) = 1 - p(x),$$

$$E(x) = \frac{\sum_x \left(\frac{S(x+1) + S(x)}{2}\right)}{S(x)}$$

These types of tables were obtained for different categories in three dimensions: geographical (33 categories: national and for every state), economic (4 categories: global and for every economic activity sector) and by the size of the business in terms of its employed personnel (13 categories: global and for each stratum of employed personnel). This way, 1,716 Survival and Mortality Tables were calculated; the image below shows the Survival and Mortality Table, calculated at national level (covering all economic activity sectors and all strata of personnel employed):

Tabla de supervivencia y mortalidad de los negocios a nivel Nacional

(x)	S(x)	p(x)	q(x)	E(x)
0	100,000	0.6672	0.3328	7.8
1	66,718	0.4794	0.5206	8.2
2	47,944	0.4209	0.5791	8.6
3	42,089	0.3793	0.6207	9
4	37,935	0.3471	0.6529	9.4
5	34,712	0.3208	0.6792	9.9
6	32,080	0.2985	0.7015	10.4
7	29,854	0.2793	0.7207	10.9
8	27,925	0.2622	0.7378	11.4
9	26,224	0.247	0.753	11.9
10	24,703	0.2333	0.7667	12.5
11	23,327	0.2207	0.7793	13.1
12	22,070	0.2091	0.7909	13.7
13	20,914	0.1984	0.8016	14.3
14	19,844	0.1885	0.8115	15
15	18,848	0.1792	0.8208	15.7
16	17,916	0.1704	0.8296	16.5
17	17,041	0.1622	0.8378	17.2
18	16,215	0.1543	0.8457	18.1
19	15,434	0.1469	0.8531	18.9
20	14,694	0.1399	0.8601	19.8
21	13,989	0.1332	0.8668	20.7
22	13,317	0.1268	0.8732	21.7
23	12,676	0.1206	0.8794	22.7
24	12,061	0.1147	0.8853	23.8
25	11,472	0.1091	0.8909	24.9
26	10,905	0.1091	0.8909	26
27	10,905	0.1091	0.8909	26
28	10,905	0.1091	0.8909	26
29	10,905	0.1091	0.8909	26
30	10,905	0.1091	0.8909	26

Fuente: INEGI. Censos Económicos 1989, 1994, 1999, 2004, 2009 y 2014.

It is important to note that a study similar to the one described here was performed considering 5 censuses, right after the 2009 census, in which 21 generations were monitored; then it was extended to include data from the 2014 census, obtaining very consistent survival, death and life expectancy probabilities, which gives more validity to the study.

Some findings resulting from these Survival and Mortality Tables are as follows:

a) In all Survival and Mortality Tables it is observed that as age increases, the number of deaths decreases, while life expectancy increases. The survival and death of businesses have an inverse behavior to human populations.

b) Small businesses die faster than the large ones.

c) Survival has different behaviors per economic activity. For example, the manufacturing sector has higher life expectancy; trade is the one with the lowest life expectancy.

d) Life expectancy, from the economic sector´s perspective, grows according to the "antiquity" of the businesses; likewise, it is different within each of them.

e) As of certain age, businesses are increasingly likely to survive and stabilize over time; this observed stability is set at earlier ages as businesses grow larger.

The complete results, as well as more details related to the methodology of this study can be consulted in the following link:
http://www.inegi.org.mx/inegi/contenidos/investigacion/Experimentales/Esperanza/default.aspx

5 Current and future work

a) More detailed indicators are being calculated at state level and also for some specific economic activities.

b) With the results derived from every census, the Survival and Mortality Tables are updated, and as the Statistical Business Register is consolidated, demographic indicators will be produced based on it.

II. South Korea's example: The Case for Korean Business Demography Statistics

South Korea has compiled and published Business Demography Statistics since 2012. South Korea has generally followed the production methods suggested by the Eurostat-OECD Manual on BDS and the SBR Guidelines with some adjustments to take into account special circumstances in Korea.

This section presents an overview of the production of South Korean Business Demography Statistics in terms of the collection of basic data, linking data, determining population, computing provisional birth and death of enterprises and confirmation of demographic events.

1 The Collection of Basic Data

Statistics Korea (KOSTAT) produces Business Demography Statistics by linking various kinds of administrative data. Also, Korea has been developing a Business Register since 2016 based on which Business Demography Statistics will be compiled in the future.

The administrative data being used for Business Demography Statistics is provided by relevant government agencies including the National Tax Service (NTS) and the Court of Korea. The NTS provides the monthly records of business registration in Korea to KOSTAT as well as the yearly records of value-added tax, the closure of business, the registration of incorporation, merger, and split off, etc.

< Administrative Data for the Business Demography Statistics >

Source	Title of Admin. Data	Data Items	Usage
		Business registration no., incorporation registration no.	Linkage key
		Name of business, name of representative, address	Testing the continuity rules
		PIN of representative	Disaggregation by gender and age
		Code of the type of main business, admin. district code	Disaggregation by industry type and region
		Establishment date	Verifying business entries
	Value-added Tax	Revenue (tax base, tax exempted business)	Verifying revenues
	Payment Statement	number of employees	Verifying number of employees
	Closure of Business	Date of closure	Verifying business exits

Source	Title of Admin. Data	Data Items	Usage
Court of Justice	Merger/Split off	Merging entity and merged entity, survived entity and newly established entity by split off	Verifying the birth and death of businesses (change of governance structure)
	Registration of Incorporation	Type of establishment	Verifying the birth and death of businesses (conversion into an incorporation)

2 Linking Data

KOSTAT currently uses the 10 digit Business Registration Number (BRN) and Incorporation Registration Numbers (13 digits) for the identification of statistical units (an enterprise) and data linkages. The original administrative purpose of the business registration numbers is for taxation, and that of the IRNs is for record registering.

3 Determining the True Population

The target of the Korean Business Demography Statistics is the entire population of active businesses in Korea, including both employers and non-employers. An active enterprise is defined to be an enterprise that has either any amount of business revenue or has more than 0 number of employee. Based on about 8 million records of the business registration data of the NTS, KOSTAT makes linkages between a parent-enterprise and its subsidiaries which have the same BRN. Non-profit enterprises and any enterprises categorized as A, O, T, U, 64992 for its industry classification are then excluded from the true population of active profit businesses. The BRN is used to identify non-profit enterprises since they have special code in their BRN.

In 2015, there are about 5.55million active profit enterprises in Korea among which 10% or 500,000 are incorporations and 90% or 5million are privately owned enterprises.

KOSTAT is considering a compilation of the Business Demography Statistics based on the population of Employer Enterprises and that of Economic Enterprises, according to the Eurostat-OECD Manual on BDS which uses employee threshold to define the scope of the population of active enterprises.

4 Screening provisional birth/death of enterprises

We compare the BRNs within the true population of active profit enterprises by year to screen provisional birth and death enterprises.

The provisional birth enterprises can be identified by comparing the BRNs in the population of active profit enterprises in year (t) with those in year (t-1). The provisional birth enterprises are identified as enterprises whose BRNs are only present in year (t). The provisional death enterprises are identified as enterprises whose BRNs are only present in year (t) and not in year (t+1).

We also use the Continuity Rules to screen provisional birth and death enterprises. An enterprise is considered as a new enterprise if more than 2 of the three continuity factors are changed in year (t) compared with those in year(t-1). Those with changes in year(t+1) compared with (t) are considered as closed enterprises.

The business sector in Korea is characterised by frequent entry and exit by businesses due to the large share of privately owned businesses. Considering this characteristic, we apply (t±1) approach in terms of the time dimension.

5 Confirmation of Demographic Events

Based on the screening of provisional birth and death enterprises, further steps are taken to identify the real birth and death of enterprises according to the Eurostat-OECD Manual on BDS and the SBR Guidelines. These manuals suggest eliminating merger, split-off, the change of ownership, break up, take-over, etc. from the population of the provisional birth and death enterprises to result in the real birth and death of enterprises. However, under the (t±1) approach used in KOSTAT, we consider an enterprise dormant for more than 1 year to be either a birth or death enterprise.

[Reference] Business Demography Statistics in 2015

- (Active Enterprises) In 2015, the number of active enterprises* was 5,554 thousand, which dropped by 5 thousand (0.1 percent) from 2014.
  - The number of persons employed in active enterprises grew by 120 thousand (0.7 percent) to 18,553 thousand in 2015.
- (Birth Enterprises) In 2015, the number of birth enterprises was 813 thousand, which declined by 30 thousand from 2014.
  - The number of persons employed in birth enterprises fell by 102 thousand to 1,244 thousand in 2015.

  * (Birth Rate) In 2015, the birth rate of enterprises stood at 14.6 percent, down 0.6%p from 2014.

  * Birth rate = (Number of birth enterprises / Number of active enterprises) x 100.

- (Death Enterprises) In 2014, the number of death enterprises was 777 thousand, which rose by 112 thousand from 2013.
  - The number of persons employed in death enterprises grew by 39 thousand to 1,009 thousand in 2014.

  * (Death Rate) In 2014, the death rate of enterprises stood at 14.0 percent, up 1.6%p from 2013.

  * Death rate = (Number of death enterprises / Number of active enterprises) x 100.

- (Survival Rate) The one-year survival rate for enterprises born in 2013 and still active in 2014 recorded 62.4 percent, rising by 2.3%p from 2013. The five-year survival rate for enterprises born in 2009 and still active in 2014 recorded 27.3 percent, dropping by 1.7%p from 2013.
- (High-growth Enterprises) In 2015, the number of high-growth enterprises was 4,077 enterprises, which dropped by 186 enterprises (4.4 percent) from 2014.
- (Gazelles Enterprises) In 2015, the number of gazelles fell by 36 enterprises (3.4 percent) to 1,024 enterprises.

III. United States' example: Annual Survey of Entrepreneurs

1 Purpose of the Study

The Annual Survey of Entrepreneurs (ASE) collects race, gender, ethnicity, and veteran status of business owners in the United States. The ASE covers firms with paid employees and is conducted on a company or firm basis rather than on an establishment basis. The data are compiled by combining survey responses with available administrative records.

2 Scope and Limitations

The ASE uses the Census Bureau's Business Register to determine cases eligible for the survey and are classified by the North American Industry Classification System (NAICS). The ASE includes U.S.-based, nonfarm businesses except those classified as:

- Crop and Animal Production (NAICS 111, 112)
- Rail Transportation (NAICS 482)
- Postal Service (NAICS 491)
- Monetary Authorities – Central Bank (NAICS 521)
- Funds, Trusts, and Other Financial Vehicles (NAICS 525)
- Religious, Grantmaking, Civic, Professional, and Similar Organizations (NAICS 813)
- Private Households (NAICS 814)
- Public Administration (NAICS 92)

3 Inputs and Requirements

The ASE uses the following inputs to determine the sampling frame:

- The Census Bureau's Business Register.
- Administrative data from the Social Security Administration.
- Lists of minority- and women-owned businesses published in syndicated magazines, located on the Internet, or disseminated by trade or special interest groups.
- Word strings in the company name indicating possible minority ownership.
- Racial distributions for various state-industry classes and racial distributions for various ZIP codes.
- Gender, ethnicity, race, and veteran status responses of a single owner business to a previous Survey of Business Owners, ASE, or the 2010 Decennial Census.

Subsequent to the survey data collection, the ASE uses the County Business Patterns and the Business Register to obtain receipts, payroll, employment, industry classification, and geography at an establishment level.

4 Methodology

Prior to data collection, each firm is assigned a likelihood for each race, Hispanic ethnicity, and gender. The likelihoods are used to assign each firm to one of nine frames:

- American Indian and Alaska Native (AIAN)
- Asian
- Black or African American
- Hispanic
- Native Hawaiian and Other Pacific Islander (NHOPI)
- Non-Hispanic, white men
- Other
- Public or otherwise unclassifiable by race, gender, or ethnicity
- Women

Each company is assigned to a single frame regardless of the potential for response in multiple categories. If a company has equal likelihood of belonging to multiple frames, then the frame that generates the highest probability of selection will be used.

The frame, metropolitan statistical area (MSA), and age of business form the strata used to select the ASE sample. Certainty cutoffs vary by sampling stratum, and sampling occurs at different rates by stratum depending on the number of sampling units available and target variances.

The ASE began with the collection of 2014 data and continued in survey years 2015 and 2016. Therefore, the MSAs used were the top 50 MSAs based on 2014 population estimates. Additionally, the remainders of states were used to capture companies located outside of the top 50 MSAs for 49 states (the Providence, RI MSA included all of the state of Rhode Island). A category of "multi-state" was used to identify companies that operated in more than one state or MSA.

The ASE uses hot deck imputation to account for non-response. Estimates are created with the Horvitz-Thompson estimator, and variance estimates are made using the delete-a-group jackknife variance estimation.

Companies are tabulated according to their responses for the main demographic categories and according to each establishment's NAICS and geography that are available from the County Business Patterns (CBP) and the Business Register. Note that this becomes a combination of firm-level data (survey responses) and establishment-level data (CBP and Business Register). For example, a company operating an establishment in New York City and another in Los Angeles would be counted once in each of those MSAs, but only once at the U.S. level. Similarly, a company operating an establishment engaged in manufacturing and another in retail trade would be counted once in each of those sectors, but only one at the U.S. level.

The company's race/gender/ethnicity/veteran status are determined by the percentage of the owners. Please note that the frame assigned during sampling no longer applies to tabulation. Tabulation is based on response data. For example, a company with a single owner owning 100% reporting both white and black for race would be tabulated as both White-Owned and Black-Owned. Each category requires more than 50% ownership to assign the company's ownership. For example, a company with three owners may report owner 1 (34%) as white, owner 2 (33%) as black, and owner 3 (33%) as American Indian (AIAN). This company would not be tabulated in any of the race categories, but would tabulate as minority-owned since 66% of the company would be classified as minority-owned.

5 Description of Results Obtained

Estimates for employer businesses include the number of firms, sales and receipts, annual payroll, and employment by gender, race, ethnicity, and veteran status. Estimates are available for the United States, each state, and the top 50 MSAs. Estimates are also available by 2-digit NAICS (also referred to as sector), years in business, receipts size of firm, and employment size of firm. Data are also collected

and presented for a variety of characteristics of businesses and characteristics of business owners. The characteristics of businesses include sources of financing, for example. The characteristics of business owners include age of the business owner, for example.

The following table provides some high level data from the 2014 ASE:

	Number of firms	Receipts ($million)
All firms	5,437,782	33,036,935
Female-owned	1,057,254	1,321,154
Hispanic-owned	298,563	335,161
Minority-owned	949,318	1,089,710
Veteran-owned	405,235	924,068

The following table shows how firms are distributed by years in business:

Years in business	Number of firms	Receipts ($million)
less than 2 years	481,981	438, 363
2 to 3 years	723,679	653,333
4 to 5 years	519,712	567,216
6 to 10 years	1,146,177	1,645,418
11 to 15 years	2,398,315	6,609,302
16 or more years	167,917	23,123,303

For more information, please visit https://www.census.gov/programs-surveys/ase.html.

Chapter 6: Business demography statistics

This Chapter identifies key indicators of business demography and provides examples of business demography statistics produced by statistics offices drawn from official statistical publications. The Chapter does not present the concepts and definitions of business demography, as these are introduced in Chapter 2 of these *Guidelines*.

6.1 Core business demography statistics

Business demography statistics support the analysis of a broad range of questions that are of high interest for research and policy. They assist the monitoring of entrepreneurial activities; the assessment of the contribution of entrepreneurship to achieving social and economic objectives; and the evaluation of the effectiveness of entrepreneurship policies.

In the economic literature on entrepreneurship and government reports on the state of entrepreneurship a set of business demography statistics emerge as most relevant for the analysis of entrepreneurial dynamics. They include the following:

- Business birth rate
- Business death rate
- Churn rate of businesses
- Business survival rates
- Share of employment creation by business births
- Share of employment destruction by business deaths
- Share of young businesses
- Employment share of young businesses

These statistics are annual and at the national level. In the past decade, these statistics have become part of the regular statistical production of NSOs in many countries; often, also breakdowns by activity sector, geographical region and other stratification characteristics (e.g. size class, legal form) are available.

Examples of production of the core business demography statistics introduced above are presented in Boxes 6.1 to 6.3: they illustrate cases where specific country concepts, definitions and/or methodology for the compilation of business demography statistics are used, as well as a case where an internationally harmonised methodology is implemented.

Box 6.1. **Businesses counts, entries and exits: Example from Australia**

Australia: The Australian Bureau of Statistics produces on an annual basis a release of counts of Australian businesses, including entries and exits, where data are sourced from the Australian Bureau of Statistics Business Register (ABSBR). The release contains counts of actively trading businesses, rates of entry to and exit from the market sector of the Australian economy, and rates of business survival. Data are disaggregated by industry, main state of operation, type of legal organisation, institutional sector, employment size range and annual turnover size range.

Source:
http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/8165.0Main%20Features1Jun%202012%20to%20Jun%202016?opendocument&tabname=Summary&prodno=8165.0&issue=Jun%202012%20to%20Jun%202016&num=&view

Box 6.2. **Birth, death and survival of enterprises: Example from the United Kingdom**

The Office for National Statistics of the United Kingdom produces an annual release on business demography statistics. The data are sourced from the Inter-Departmental Business Register (IDBR), while the guidelines found in the *Eurostat-OECD Manual on Business Demography* are used to produce the business demography statistics. The release presents counts and rates of enterprise births and deaths as well as survival rates, with breakdowns by broad industry group and by UK region.

Source:
https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/bulletins/businessdemography/previousReleases

Box 6.3. **Young firms and employment creation: Example from the United States**

The BLS Business Employment Dynamics (BED) programme of the U.S. Bureau of Labor Statistics (BLS) produces data series on new businesses and job creation. This set of statistics is generated from the Quarterly Census of Employment and Wages programme. The BED data series on age of establishment tracks cohorts of new business establishments born in the same year and reports on their associated employment.

Source: https://www.bls.gov/bdm/entrepreneurship/entrepreneurship.htm

6.2 Demography statistics on businesses with high-growth

In the context of the analysis of business dynamisms and entrepreneurial phenomena, the role of firms with high growth as drivers of job and wealth creation has attracted the interest of policy makers. Official statistics on high-growth enterprises are produced in particular in the Member States of the European Union and in countries contributing to the OECD-Eurostat Entrepreneurship Indicators Programme. The relevant indicators include:

- Rate of high-growth enterprises
- Employment in high-growth enterprises

Indicators that focus on a subset of high-growth enterprises, i.e. those younger than a defined age, the "gazelles", are also of interest for economic analysis and are compiled in some countries.

The variable used for measuring growth is either employment or turnover, although high-growth measures based on employment growth are more common (see Chapter 2 for concepts and definitions).

Box 6.4. presents the example of business demography statistics on high-growth enterprises produced by the Member States of the European Union.

Box 6.4. **High-growth enterprises: Example from the European Union**

Eurostat regularly issues news releases to disseminate business demography data collected from Member States on the basis of the recommendations of the *Eurostat-OECD Manual on Business Demography Statistics*.

A news release of October 2016 presents a selection of the data available on high-growth enterprises in the European Union. Data come from the Eurostat Business demography statistics, which cover variables on characteristics and demography of the business population, including information on high-growth enterprises. Data are mainly drawn from business registers. A high-growth enterprise is defined as an enterprise with average annualised growth in number of employees greater than 10% per year over a three-year period and having at least 10 employees at the beginning of the growth.

Source: http://ec.europa.eu/eurostat/documents/2995521/7706167/4-26102016-AP-EN.pdf/20f0c515-ed43-45c3-ad6a-ca0b26b36de5

6.3 New indicators from micro data linking

As explained in Chapter 4 of these *Guidelines*, linking at the micro level data from the Business Register to administrative data and other datasets or survey data on businesses as well as on individuals has provided new insights into the relationship between the performance of an enterprise and the characteristics of enterprise and of its entrepreneur(s). While these developments are relatively new, there are already many successful examples of new statistics generated from the micro linking of data. Boxes 6.5 and 6.6 introduce two relevant examples of linking business demography statistics respectively with other business data and with data on individuals.

6.3.1 New indicators: Business demography statistics linked with data on individuals

Indicators on the profile of the entrepreneur are produced from the linking of business demography data with data on individuals (see Chapter 4 for methodologies of data linking). Differently from statistics on self-employed that can only describe the characteristics of the entrepreneur but not of the business he/she owns, linked data permit investigating business dynamisms by gender of the business owner. Box 6.5 provides an example of official statistics on the profile of the entrepreneur.

Box 6.5. **The profile of the entrepreneur: Example from Italy**

ISTAT has created new indicators to identify the main profiles of new entrepreneurs in Italy. The linking methodology implemented to produce the indicators is consistent with the recommendations of the *Eurostat-OECD Manual on Business Demography Statistics*.

The methodology developed to profile the entrepreneur exploits information of administrative sources and official statistics. In particular, the basic informative structure is represented by an integrated system of administrative files on occupations (of LEED type, i.e Linked Employer Employees Database) that allows connecting each individual with the enterprise where he/she performs an activity. The identification of an entrepreneur within an enterprise is done through the application of different deterministic rules according to the juridical form of enterprises. This informative structure is enriched with demographic information on the enterprise.

A statistical release prepared by ISTAT in December 2016 presents the methodology and the new indicators.

Source: https://www.istat.it/it/archivio/194762

Chapter 7: Topics for further work and research

7.1 Introduction

The topics dealt with in these Guidelines are to a considerable extent quite new ones: only few years ago statistics on business demography was developed. The same is true for using linking methodology between different databases to achieve new kinds of statistics without any additional burden on enterprises. The Guidelines aim at providing methodological and practical guidance to countries for establishing business demography and entrepreneurship statistics with the focus on the use of SBRs as the basic database. Given the status of these approaches, the Guidelines could not cover all relevant issues in an appropriate manner. Furthermore, due to the current implementation of these methodologies in many countries new methodologies and approaches will emerge.

Share of experience between countries and common development projects should be encouraged as developing new methodologies and good practices is resource demanding. International cooperation can help to reduce the gaps between developed and less developed statistical systems, while capacity building and training activities will be useful to support countries with less experience in these fields. Sharing of methods and practices may help improve international comparability of business demography and entrepreneurship statistics.

During the work of the drafting of the Guidelines, the Task Force noted a number of topics where further work and research is recommended. These topics are listed and briefly described below.

7.2 Recommended topics by domain

7.2.1 Requirements of the SBR system to support business demography statistics

The focus of the Guidelines are the requirements of the SBR system to support business demography statistics, namely to creation of a longitudinal database that provides the long-term linkages between the business units in the SBR. Such a database is needed as the focus in the maintenance of the SBR is the on the up-to-date population of the businesses for the provision of survey frames. The Guidelines provide guidance on the creation of such a database and discusses ways of the integration into the SBR system. The SBR system needs to be enlarged to support the longitudinal aspects. For the creation of the longitudinal data record linkages methods might be required to identify the continued units in the SBR. These methods are also required for the maintenance of the SBRs, for instance to link with administrative sources.

Work on these issues will be needed for the national implementations; however, exchange of experience and common projects will gain additional insights into these issues and will help other countries too:

Methodology to create longitudinal databases of the units in the SBR.

Conceptual enlargement of the SBR system to support the integration of the longitudinal aspects.

Experience with record linkage methods.

7.2.2 Framework of business demography statistics

Business demography statistics provide data on the dynamics of the business population over time and core variables that describe the development of the business population. Focus of business demography statistics is on the birth of new businesses, the survival of newly created businesses and the death of businesses. Based on kinds of data various indicators can be derived, such as birth, survival and death rates. There are also different concepts of business demography statistics, such as overall business demography or employer business demography. Furthermore, different statistical units can be used. The focus may be on national or regional data, or on annual or sub-annual frequencies. A crucial issue in business demography is the definition of the entrepreneur. Definitions could be based

on legal or administrative basis (e.g. self-employed persons) or on an economic basis (persons that are leading and managing the businesses, irrespective of the legal status).

Actual implementation in the countries may establish statistics using one or the other business demography concept. Due to the lack of comprehensive implementation, not all concepts are equally developed and tested. Topics for further work could be:

- Analysis and improved specification of the basic concepts and definitions of business demography.
- Analysis and testing of regional business demography.
- Analysis and testing of sub-annual business demography.
- Comparative analysis of business demography using different statistical units.
- Analysis of the different concepts and definitions of entrepreneurs and possibilities of implementation.

7.2.3 Survey approaches to compile business demography data

Due to various reasons, some countries are compiling business demography statistics based on survey approaches rather than on the SBR. Such approaches are usually quite expensive and may not be repeated on a regular basis. It might also not be fully possible to apply the international concepts and definitions due to the restrictions given by a survey. However, survey approaches may be a step towards regular data compilation using the SBRs. Work on these issues could cover:

- Analysis of the deviations of survey based approaches from the international concepts and definitions.
- Approaches to integrate survey based data into the SBR system.
- Strategies for switching from survey based approaches to SBR based approaches.

7.2.4 Use of administrative sources

Administrative sources are the main data inputs for the maintenance and update of the SBRs. Their structure and quality determine the possibilities of the actual maintenance and the quality of the survey frames provided from the SBR. Administrative sources are thus also central for the creation of the longitudinal databases that could serve as the basis for business demography statistics.

The cooperation with the administrative data providers is thus crucial for the SBRs. With respect to the establishment of business demography statistics a topic for further work in the countries could be:

- Analysis of the available administrative data in the country with respect to the requirements for establishing business demography statistics and proposals for their improvement to serve the SBR work.

7.2.5 Linking of databases to achieve additional characteristics of the businesses

SBRs cover only a restricted set of variables for the each of the business units. Having characterised each business unit according to the demography events, will not increase the variables available. Thus, linking the business units with data from other business or social statistics will enlarge the analytical possibilities and will gain new insights into the dynamics of business development.

- Linking of databases on micro level has become a quite important approach, as enlarged databases are established without additional burden on the businesses. Work on the following issues should be encouraged:
- Linking methodology in case of missing a common identifier and in cases where one of the liked databases cover only a sample of the population.
- Approaches of linking business demography data with other business data.
- Approaches of linking business demography data with social data on the entrepreneurs and on persons employed.

7.2.6 International comparability

International comparability of business demography and entrepreneurship statistics is of high importance. Countries are requested to implement international methods and concepts as far as possible. The Guidelines supports this goal by providing guidance on concepts, definitions and methods, and by providing some country examples. Any work in this area will improve international comparability:

Analysis of how the national implementation is diverging from international standards and developing of strategies in order to increase compliance.

In case that a country does not yet produce business demography statistics or only some aspects of it, work on an appropriate planning and developing of the implementation methods and procedures should be undertaken.

[i] See: http://ec.europa.eu/eurostat/web/structural-business-statistics/overview.

[ii] See: http://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/KS-RA-07-015.

[iii] Intrastat regulation: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32004R0638&from=EN

[iv] See: http://ec.europa.eu/eurostat/web/international-trade-in-goods/data/focus-on-enterprise-characteristics-tec

[v] More details are available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Services_trade_by_enterprise_characteristics_-_STEC

[vi] http://ec.europa.eu/eurostat/web/international-trade-in-goods/data/focus-on-enterprise-characteristics-tec

[vii] http://ec.europa.eu/eurostat/documents/3859598/5922981/KS-RA-12-016-EN.PDF/c93cdf48-5efa-459f-b218-731a9a5476e9?version=1.0

[viii] Resolution concerning the International Classification of Status in Employment (ICSE), adopted by the Fifteenth International Conference of Labour Statisticians (January 1993) http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_087562.pdf

[ix] https://unstats.un.org/unsd/nationalaccount/sna2008.asp

[x] de Waal, T. (2015). General Approaches for Consistent Estimation based on Administrative Data and Surveys. Discussion paper 11, Statistics Netherlands.

[xi] Kroese, A.H. and R.H. Renssen (2000), New Applications of Old Weighting Techniques; Constructing a Consistent Set of Estimates Based on Data from Different surveys. In: Proceedings of ICES II. American Statistical Association, Buffalo NY, pp. 831-840.

[xii] Shlomo, N., T. de Waal and J. Pannekoek (2009), Mass Imputation for Building a Numerical Statistical Database. UN/ECE Work Session on Statistical Data Editing, Neuchâtel, Switzerland.