

## Chapter 7. Maintenance of SBR

### INTRODUCTION

This chapter discusses the maintenance of the SBR. Key elements for the maintenance of a SBR are the sources of information about statistical units, the main input coming from an administrative source. The SBR should be partitioned into maintenance groups, allowing different maintenance procedures according to the size and complexity of the enterprises. Finally, maintenance of the SBR should be considered in relationship with frame construction and with the system of economic surveys based on the SBR. In this regard, the time dimension plays an important role.

The quality issues pertaining to the maintenance of the SBR are presented in Chapter 8 and the IT aspects are treated in Chapter 9.

In Section 7.1 different aspects of the maintenance strategy for a SBR are presented, i.e. the sources of information, maintenance groups and the time dimension. In Section 7.2 frame construction in relationship with the maintenance processes of the SBR is taken up. Section 7.3 offers a summary of the different types of demographic events pertaining to statistical units. In Section 7.4 the updating procedures are discussed. Certain quality management aspects are briefly reviewed in Section 7.5 and finally Section 7.6 is about the treatment of errors.

The references for this Chapter are

Eurostat, 2010, Business Registers Recommendations Manual  
Chapter 12: A general overview of demographic events,  
Chapter 13: Demographic changes concerning the enterprise,  
Chapter 14: Continuity rules for the enterprise,  
Chapter 15: Demographic changes concerning the local unit, and  
Chapter 16: Continuity rules for the local unit of EUROSTAT,

African Development Bank, 2012, Guidelines for Building Statistical Business Registers in Africa  
Chapter 10: Specification of SBR Maintenance Strategy and Procedures

Eurostat – OECD, 2007, Manual on Business Demography Statistic  
Chapter 4: Typology of Demographic Events

## 7.1. SBR MAINTENANCE STRATEGY

Maintenance is a constant update of the SBR in terms of both *coverage* and *content*. The units have to be identified, with no over-coverage or under-coverage and no duplicates. The information about the units must be available and sufficiently detailed to define the statistical units and to determine their activity classification. The information should be collected just once. The dates of changes should be known.

A country's economy is constantly changing – new businesses are formed, existing businesses merge, change production activities or location, go bankrupt, etc. To ensure that the SBR enterprises (and other standard statistical units) remain aligned with and representative of legal units and their productive activities, these changes have to be detected and the SBR has to be correspondingly updated. This process as is referred to as *SBR maintenance*. (African Development Bank, 2012, Guidelines for Building Statistical Business Registers in Africa, Chapter 5.7)

The maintenance strategy takes into account the following aspects: (1) the sources of information (administrative sources, feedback from surveys and control surveys, also considering rules for dealing with conflicting information), (2) the maintenance groups, (3) the time dimension (timing of the updates, historical register).

### 7.1.1 Administrative sources

Each government regulation relating to businesses results in an administrative business register of businesses bound by that regulation and transaction data resulting from application of the regulation.

It is recommended that only one primary administrative source is chosen which provide the best possible coverage and content. Criteria for the choice of a primary administrative source can be access, coverage, content and the actuality and the quality of the data.

If secondary administrative sources are used, care must be taken that no duplications of units arise from using several administrative sources. Duplications can be avoided if the secondary source has common unit and identification scheme with the primary source, or contains only a very small numbers of units (check by manual investigation) or has no overlap with the primary source.

### 7.1.2 Feedback from surveys

Feedback of frame data obtained from previously conducted economic surveys is a potential source of information about enterprises and (depending upon the survey) establishments. By design, the first few questions asked by any business survey should constitute a check of the data items in the survey frame, such as name, address, contact information, and activity status. Subsequent questions may collect updated versions of economic activity and size measures.

There are no technical problems with using data for enterprises that have been sampled with certainty, provided the response rate is close to one. However, for medium-size and small enterprises that are sampled with probability less than one, there is a potential for causing bias in future survey samples. It is therefore recommended that feedback from surveys about attributes used for sampling for enterprises that were not selected with certainty should not be used to update the corresponding enterprise attribute values in the SBR. The information can be used as a signal to trigger further maintenance operations.

### 7.1.3 Control surveys

Objectives of a control survey are to verify existing data item values and to obtain the values of missing data items.

Control surveys are conducted on a continuous basis, with regular selection of a new sample. There is no focus on estimation.

The total sample size determined by taking into account the frequency with which changes occur in the economic world, the quality of data received from the administrative sources, the quality of data in the SBR and the resources available.

### 7.1.4 Dealing with conflicting information

The SBR can be updated from a wide range of different sources: administrative sources, profiling, feedback from surveys, control surveys.

Several sources may provide values for register characteristics and where this happens there will always be a degree of conflict between the sources. This raises the question of which source to use.

The best way to answer this question is to start gaining a thorough understanding of each source, covering issues such as the methods of data collection and validation, the time of collection and the relative importance assigned to that variable by the source. This understanding then allows the sources to be prioritized, either for all units or for different categories of units depending on one or more other criteria.

Once priorities have been determined for a certain characteristic, the next step is to devise a way to apply them to register updating procedures. This can be done in several ways. Perhaps the most reliable method is to store a source code and date with the characteristic and to implement algorithms to determine which combinations of source code and date can overwrite another. The date is important here, because it may be decided that data from a lower priority source should overwrite that from what would normally be seen as a better source if it is significantly more recent.

### 7.1.5 Maintenance groups

The amount of maintenance effort devoted to an enterprise should be in accordance with its size and potential impact upon published statistics, and should take into account its propensity to change and the sources of updating information. Thus, enterprises should be partitioned by size and potential impact into maintenance groups, each of which is subject to a particular set of updating procedures.

#### ***Swiss Business Register: maintenance procedures***

For maintenance purposes, the enterprises are divided into three groups.

- Enterprises subject to profiling.  
All enterprises with more than 10 local units or more than 100 employees are integrated in the group profiling. The profiling staff contacts them directly each 3 month, in order to gather structure information and employment data.
- Enterprises subject to profiling light.  
All other enterprises with less than 10 local units or less than 100 employees are integrated in the group profiling light. The profiling staff contacts them each once a year by internet survey in order to gather structure information and employment data.

- All enterprises with only one local unit, including small businesses. These are maintained using updating information from administrative sources and business survey feedback.

This maintenance strategy is dictated by the necessity to determine the employment distribution in all the local units. The regional distribution of employment is an essential information of the Swiss business statistics and this drives the maintenance strategy. All enterprises with only one local unit are integrated in the control survey of activity code that allows a quality control of the activity code for the whole SBR.

#### ***Statistics South Africa Business Register: maintenance procedures***

For maintenance purposes, enterprises are divided into two groups.

- Enterprises subject to profiling.  
This set of enterprises is maintained by profiling operations. Indications of changes from administrative sources or from business surveys are used as signals indicating the need for re-profiling. The enterprises are not included in SBR control surveys, nor are they subject to ad hoc investigations.
- All other enterprises.  
All other enterprises are maintained using updating information from administrative sources and business survey feedback, which is supplemented as needed by information gathered by SBR control surveys and ad hoc investigations.

#### ***Australian Bureau of Statistics (ABS) Business Register: maintenance procedures***

For maintenance purposes, enterprises are divided into groups.

- ABS maintained enterprises.  
These enterprises are maintained by the SBR staff by profiling, BR survey, business survey feedback and ad hoc investigations, as appropriate according to their size.
- Australian Tax Office (ATO) maintained enterprise.  
These enterprises are maintained entirely by taxation data from the ATO. No ABS sources are used. Even if evidence of a change to an enterprise is detected by the ABS, for example through a business survey, BR data for the enterprise are not updated unless and until the change manifests itself through the taxation source. The benefits of expending no ABS maintenance effort are considered to outweigh the failure to make use of the ABS derived information.

### **7.1.6 Timing of the updates**

The register must be updated at least annually to record unit creations and deletions, as well as changes in address and stratification variables.

The timing of the updates must be made in a systematic and controlled way, taking into account the economic statistics program, in particular the conflicting demands from structural and short term surveys (up-to-date information vs. stability and coherence). Further the updates to the SBR must be coordinated with construction process of the sampling frames.

It is also important that the users of the register are informed about the calendar of the updates of the SBR.

### **7.1.7 Historical register**

An historical register is a functionality of the SBR allowing the reconstruction of the history of the units. As a prerequisite, the dates of the demographic events must be recorded.

With this it would become possible

- to obtain information about survival and continuity of units
- to obtain information about size development (how fast a unit or group of units is growing, through what kind of economic activity)
- to analyze changes in business organizations, like for example location, judicial and financial links and in economic activity
- to reconstruct the state of the register for any given date or any reference period

The historical register could be maintained in three different ways

- A new register is put in place at the beginning of each year. During the year, all changes that a unit undergoes and the dates and reasons for those changes would be recorded for each unit. At the end of the year, the register would be copied and stored.
- The register can be analyzed at regular intervals (e.g. annually) to provide a series of pictures of the structure of the register, e.g. the numbers of legal units and statistical units by type, economic activity, employment and turnover size groups, institutional sector, region, etc.
- The register stores information about changes as they occur. These changes are date stamped to record the date they occurred in the register, and/or the date they occurred in the real world. This allows populations of businesses to be constructed at any point since the register started operating in this way. Storing the reason for the change is often more difficult, but basic categories of change can often be determined by automatic rules, e.g. the simultaneous transfer of all local units from two enterprises to one new enterprise is likely to reflect a merger.

Additionally, an historical register allows analyses of register data to be targeted more closely at user needs, though care must be taken to avoid confusing users with different data sets for what appears to them to be the same point in time. Key register users, particularly those with direct access to the register, must be trained so that they are aware why the historical versions exist and why they may differ for certain units at certain times.

## 7.2 SERVICES

The primary service (output) of the SBR is to supply sampling frames for all the economic surveys.

Other services may include samples, survey control files and survey shell databases for these surveys, register-based statistics (in particular enterprise demography), linkages between economic data collected by different surveys, monitoring of response burden.

### 7.2.1 Sampling frame

It is recommended to use a common enterprise frame as a basis for the creation of survey frames for specific surveys. The common enterprise frame is derived from regular snapshot of enterprises in the SBR. The common frame approach allows producing more harmonized survey frames and enables more comprehensive quality control.

#### *Creation of an Enterprise Snapshot*

An enterprise snapshot is a flat file extracted from the SBR at a given point in time, containing every enterprise in the SBR together with all the variables needed for survey sample selection: identification data (enterprise name, identification code), descriptive data (legal form, institutional sector, economic activity, size, region, activity status) and demographic data (date of creation, change of structure or activity, activity status).

Counts of enterprises by region and by economic activity sector should be compared with counts for the previous snapshot.

#### *Creation of a Common Enterprise Frame*

Selection of the enterprises that are active and that do have a sufficiently detailed data items (economic activity code, size, etc.) for frame definition purposes.

Common Enterprise Frame counts should be checked against counts for the previous period common enterprise frame.

#### *Creation of Survey Frames*

For each survey, the survey frame is extracted from the common enterprise frame by including just those enterprises that are in scope for the survey according to the survey frame specification. Frame specification include the reference period, the sampling unit (normally the enterprise), the population of units, the data item necessary for sample selection and contact data items (for the sampled units only).

The survey frame counts should be checked against values for the previous survey cycle, if there was one. Checks should be made for unexpected changes of activity status, size, or economic activity of enterprises known to make significant contributions to the survey estimates.

### 7.2.2 Update of SBR and sampling frame

Maintenance of the SBR should be considered in relationship with frame construction, and also with the program of surveys based on the SBR.

The SBR is updated from sources of information, each having a given periodicity, coverage and content. This must be taken into account to build a coherent system for SBR updating, frame construction and sample selection for the business surveys based on the SBR. Such an approach allows a better description, as a function of time, of the effective coverage and content of the SBR and derived sampling frame, an important aspect for the users of the SBR.

To illustrate some of the issues, we consider as an example a simplified model of the update of a SBR. Specifically, we consider only the population of simple enterprises and we assume that these units are updated using only one administrative source, with the following properties:

- each administrative unit is linked to only one enterprise;
- for the enterprises linked to the administrative source, death can be detected on a continuous basis from the death of the corresponding administrative unit;
- the birth of an administrative unit is potentially a new enterprise, but new administrative units are not all relevant for the SBR (possible over-coverage);
- The administrative source lags by one year.

Given the time lag of the administrative source, the table below gives the availability of data from the administrative source.

Calendar date	Reference period
2011_12	2010
2012_12	2011
2013_12	2012
2014_12	2013

For example at the end of 2013 (Calendar date = 2013\_12, i.e. December 2013) we know the population of active administrative units during the reference year 2012. We denote by  $A(2012)$  the population of active administrative units during the reference year 2012.

Potential births of enterprises can be detected by comparing the populations of active administrative units for two consecutive years. Taking again as an example the situation in December 2013, we have available the populations of administrative units  $A(2012)$  and  $A(2011)$ . Taking the difference of these two populations, we can identify the administrative units created during the reference year 2012. We denote by  $B\_admin(2012)$  the population of administrative units created during the reference year 2012.

In the situation considered here, a newly created administrative unit may not necessarily correspond to an SBR enterprise. This information must be obtained through a control survey. For the population  $B\_admin(2012)$ , this control survey takes place during 2014 and ends in June 2014, say. At the end of the control survey, we have identified the administrative units which correspond to new enterprises. We denote by  $B(2012)$  the population of enterprises created during the reference year 2012. The table below gives the availability of data from the administrative source and from the control survey identifying the new SBR enterprises.

Calendar date	Active units	Births detected	New enterprises
2011_12	A(2010)	B_admin(2010)	
2012_12	A(2011)	B_admin (2011)	B(2010)
2013_12	A(2012)	B_admin (2012)	B(2011)
2014_12	A(2013)		B(2012)

Based on this process of updating the SBR, choices can be made of when to construct new sampling frames. For example, it may be decided to create a new version of the common enterprise frame each year in December.

Taking again December 2013 as an example, the coverage would then correspond to the enterprises created up to the end of 2011, less all the enterprises which have disappeared up to the end of 2013. The enterprises created during 2012 are known to be a subset of the administrative units created during 2012 but have not been yet identified through a control survey. The enterprises created during 2013 are completely missing. For the enterprises covered, the content as from the administrative source corresponds to the reference year 2012.

## 7.3 DEMOGRAPHIC CHANGES AND CONTINUITY

A demographic event is based on changes in the existence of production factors and their distribution within and among business organizations. It also covers follow-up of units in time dimension (survival or discontinuity) and development in time according to certain characteristics like size or type of activity.

### 7.3.1 General typology of demographic events

A general typology of demographic events is based on the distinction between events involving existential changes, i.e. the emergence or disappearance of combinations of production factors and distribution changes i.e. changes in the distribution of production factors between units

<b>1</b>			<b>Changes of existence of combinations of production factors</b>
	1.1		Emergence of combinations of production factors (birth)
	1.2		Disappearance of combinations of production factors (death)
<b>2</b>			<b>Changes in the distribution of production factors</b>
	2.1		Redistribution of the production factors within one enterprise
		2.1.1	Redistribution of production factors across local units
	2.2		Redistribution of the production factors of more than one enterprise
		2.2.1	Concentration of enterprises (merger, takeover)
		2.2.2	De-concentration of enterprises
		2.2.3	Transfer of production factors between enterprises
		2.2.4	Enterprise restructuring
	2.3		Redistribution of the production factors within one enterprise group
		2.3.1	Redistribution of production factors across local units of more than one enterprise
		2.3.2	Redistribution of production factors across enterprises
	2.4		Redistribution of the production factors of more than one enterprise group
		2.4.1	Concentration of enterprise groups
		2.4.2	De-concentration of enterprise groups
		2.4.3	Transfer of production factors between enterprise groups
		2.4.4	Enterprise group restructuring

### 7.3.2 Criteria for the identification of demographic events

A demographic event is defined as an event which has an impact on the existence of statistical units and links between them.

The enterprise is the central unit of the statistical system. 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 (KAU) and local unit), or a combination of enterprises bound together by legal and/or financial links (enterprise group). Thus demographic events are identified and classified in a way that reflects the central position of the enterprise in the statistical system.

The enterprise is defined as an organisational unit producing goods or services which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. A demographic event is therefore linked to either changes in the existence of such combinations of production factors or changes in the existence of parts of those combinations of production factors, if these parts are distinct in respect of economic activity or location.

A demographic event can also occur as a result of changes in the distribution of existing production factors, if they result in changes in the existence of the statistical units involved or in the links between them. At the enterprise level this means changes at the KAU and/or local unit level.

Redistribution can also take place among enterprises. The number of enterprises involved may decrease (concentration, e.g. merger of enterprises), increase (de-concentration, e.g. split-off of an enterprise from an existing enterprise) or stay constant (KAU or local unit transfer, restructuring).

If populations of statistical units are followed over time, changes in size, economic activity and location are of interest. Changes to characteristics may also be caused by non-demographic events.

### **7.3.3 Typologies of demographic events and continuity rules for the enterprises**

#### *Demographic events*

The birth of an enterprise is the creation of a combination of production factors and a death is the dissolution of a combination of production factors, both with the restriction that no other enterprises are involved in the event.

A change of ownership is where a new legal unit is formed to take over the activities of an existing enterprise. This event does not affect the continuity of the enterprise; therefore there are no creations or deletions of enterprises.

Restructuring within an enterprise (e.g. creation of a local unit) does not affect the continuity of the enterprise. Restructuring may affect characteristics such as size or economic activity.

A concentration is an event involving more than one enterprise before and one enterprise after the event. If all the enterprises before the event lose their identity, the event is a merger; if one of them retains its identity, the event is a takeover. Mergers and takeovers do not involve the death or birth of units. As a consequence of a takeover, some characteristics (size, economic activity) of the enterprise that retains its identity may change.

A de-concentration is an event involving one enterprise before and more than one enterprise after the event. In a break-up, the enterprise is divided in such a way that none of the new enterprises retains the identity of the original enterprise. In a split-off, the identity of the original enterprise is retained by the larger enterprise.

#### *Translation of events into register updates*

Births and deaths involve creating new identity numbers or deleting existing ones respectively. Date of birth and date of death are more difficult to determine.

Changes of ownership and restructuring are reflected through changes to relationships or characteristics.

In the case of a merger, all identity numbers of the enterprises existing before the event are deleted and an identity number for the emerging enterprise is created. In the case of a takeover, the enterprise that takes over the other enterprises retains its identity number and the identity numbers of the enterprises that are taken over are deleted in the register.

In the case of a break-up, all identity numbers of the enterprises are created after the event and the identity number of the original enterprise is deleted. In the case of a split-off, a new identity number is only given to the split-off enterprises.

### *Information sources*

Most events are detected after reception of signals from the administrative sources of the statistical business register. Additional data may be needed, either directly collected or from other sources. The policy of register updating may differentiate between micro, small, medium and large enterprises.

Administrative creations and cessations do not necessarily result in births and deaths of enterprises and enterprises may be born and may die without a change in the legal unit(s) of the enterprises. Detection of enterprise deaths is more difficult than detection of enterprise births.

### *Reconstruction of demographic events from business register information*

The cause of an enterprise creation can be a birth, a merger, a break-up or a split-off. The cause of a deletion can be a death, a merger, a takeover or a break-up.

For reconstruction of a demographic event, one needs to know which enterprises were involved in the event. For the cases of concentration and de-concentration, it is necessary to register a link over time between the enterprises involved. In the case of a merger, the original enterprises must be linked to the emerging enterprise. In the case of a takeover, the enterprise which is taken over must be linked to the surviving enterprise. These links must be recorded with dates, i.e. the business register must be a historical register.

### *Definition of continuity*

An enterprise is considered to be continued if its production factors are continued. It is discontinued if its production factors are discontinued. In practice, this means the continuity of at least two of control (same controlling legal unit), economic activity and location.

## **7.3.4 Typologies of demographic events and continuity rules for the local unit**

### *Demographic events*

The birth of a local unit is the creation of a (partial) combination of production factors at a geographically identified place. A death is their dissolution.

The transfer of a local unit amounts to a change of links.

### *Reconstruction of demographic events from business register information*

Births and deaths involve creating new identity numbers for the local units or deleting existing ones respectively. Date of birth and date of death are more difficult to determine.

The transfer of a local unit is registered by the deletion of the link between the local unit and the enterprise to which it belonged before the event, and the creation of a link between the local unit and the enterprise to which it belongs after the event.

It is necessary to keep historical information, including historical links between local units and enterprises.

### *Definition of continuity*

Continuity of the local unit is defined in terms of continuity of location and production factors, with emphasis on production factors that can be readily identified at the level of the local unit (land, buildings, employment).

If the location is the same, the continuity of production factors, in particular of employment, and the continuity of the enterprise to which the local unit belongs are the criteria for continuity.

A local unit which does not change location loses its identity if at least two of the following three factors change: the enterprise identity, the principal activity (NACE at 4 digit level), and at least 50 percent of the local unit employment.

If a local unit moves within a region defined at the local level, it retains its identity if none of the three factors mentioned above change. Otherwise, it loses its identity.

If a local unit moves outside the region it is deemed not to be continued.

A local unit is continued if it resumes its seasonal activities or resumes its activities within 24 months.

The birth of an enterprise, which comprises only one local unit, implies the birth of a local unit, and the death of such an enterprise implies the death of a local unit. However, a local unit equaling an enterprise can discontinue, followed by the creation of another local unit at the same location, whereas the enterprise may remain the same.

## **7.4 THE HANDLING OF CHANGES IN THE STATISTICAL BUSINESS REGISTER**

A business is not a static entity; it constantly undergoes changes like restructuring, concentration, redistribution, death, among others. For that reason it is necessary to define policies and strategies about how to maintain SBR. The way changes are handled, can affect the consistency and compatibility of register-based statistics.

When a group of characteristics is updated, it is important to know that each group will impact the statistical process in a different way. We have to build a comprehensive list of changes that has to be monitored and treated.

### **7.4.1 Changes in identification variables**

Identification variables are basic to follow a unit over time. This is very important especially when there is conflicting information about the dates when changes happened. Furthermore identification variables are basic for quality data management.

Identification variables allow identifying a unit and its links with other register units or other sources of information. Identification variables normally depend on administrative creation or deletions rather than changes to existing data. These changes can be recorded in statistical business register after they happen or regularly several times per year.

The changes in identification characteristics should normally be reflected in statistical business registers as soon as they are notified. This is because these details are used for mailing survey forms and for geographical analyses of register data. That change to one unit does not necessarily imply changes to other linked units.

Changes to legal form may affect the consistency of survey samples and population. Crosschecking name and legal form can help identifying problems, like change of legal form from unincorporated enterprise to a limited corporation.

### **7.4.2 Change in economic/stratification variables**

These characteristics affect the probability of a unit being included in a particular survey. It is therefore important to consider whether to update these characteristics as soon as new information is received and risk increased volatility in survey samples and populations or to hold them until a certain point in the annual cycle of surveys where the impact of changes on the consistency of results would be reduced.

Ideally, all the information in SBR has to be continuously updated, but sometimes it is not always possible to do that. To make this process easier one option could be to define a group of priority units that could be updated first.

Changes to size variables (person employed, employees, employees in full-time equivalent and turnover) have a significant effect on survey samples. Therefore the operation of update must take place before a new sampling frame is constructed. Changes to geographical location code have been effected the enterprise or local unit continuity rules. Therefore the operation of update must take place before a new sampling frame is constructed.

Large and complex units are, in general, likely to be more problematic. Changes to such units have more of an impact on users or in register-based statistics, so there is more of an incentive to ensure that they are correct. The database of those units could be updated with information collected by surveys, because most units are included in the population frame for several years. Also the

maintenance could be done by comparing data from a range of sources to see if it is consistent, or by contacting the unit concerned to validate the change.

Small units are less complex and for that reason the process to maintain their databases requires less validation and makes it unnecessary to contact the unit to confirm the changes.

Information received on changes in the various activities of a particular unit may prompt a change to the principal or secondary activity code, or ancillary activity marker for that unit. These changes may be sudden, e.g. due to a change in management policy for the business concerned, or gradual, where the balance of activities has shifted over time. In such cases, it is recommended to follow a stability rule whereby the change in the principal activity should be made when the current one has been accounting for less than 50 % of the value added for at least two years. This rule helps to reduce movements in business demography, which are no more than temporary phenomena or statistical artifacts.

For businesses engaged in certain activities, e.g. tourism, there is likely to be a seasonal pattern to the number of persons employed (and possibly also to short-term turnover data). This means that data from a particular point in time could give a misleading view of the size of a business. It is therefore recommended that averages for the operational period should be held for these variables. Some users may argue, however, that this policy restricts the use of the register for detecting short-term trends, and particularly the early detection of turning points. If this is an issue, the simplest solution is to store both current values and averages.

When economic/stratification characteristics are updated, it is useful to record the date of the change, the source of the new information and the previous value. These are very valuable pieces of information to help assess the quality of variables and to audit the change process.

#### **7.4.3 Change in demographic characteristics**

Demographic events may have an impact on the characteristics of units, such as their principal economic activity or size. Links between legal units are also relevant when discussing demographic changes, because they determine the constitution of the enterprise group.

The update of these changes would affect the probability of a unit being included in a particular survey. Therefore it is recommendable to hold them until a certain point in the annual cycle of surveys where the impact of changes on the consistency of results would be reduced.

#### **7.4.4 SBR Updating Procedures**

##### *Legal and operational units*

Information about legal and operational units is obtained through profiling. A new profile corresponds to the creation of a new set of legal and organizational units. Re-profiling means updating an existing set of legal and organizational units and their attributes.

##### *Administrative units*

Information about administrative units is obtained from the corresponding administrative sources

- In the case of a new administrative unit appearing in the incoming administrative data, a corresponding new administrative unit must be created in the SBR administrative data.
- In the case of changes in attribute values of administrative units appearing in the incoming administrative data, the values in the corresponding administrative units in the SBR administrative data must be updated.

- In the case of an administrative unit being marked as canceled, or simply not appearing in the incoming administrative data, the activity status of corresponding administrative unit in the SBR administrative data must be updated accordingly, for example changed to dead.

#### *Statistical units*

The updating procedures for statistical depends on defining the relationships associating legal, operational, and administrative unit changes with statistical unit updates.

The specification of the updating procedure is based on a matrix with:

- one axis containing all possible types of legal and administrative signals that can be detected
- the second axis containing the various types of statistical unit change
- the cell entries defining the rules regarding how each signal is to be reflected in appropriate updates to the statistical units, taking into account continuity rules, stability rules, and rules for dealing with conflicting information

#### *Statistical unit birth, death, and continuity rules*

Updating procedures must incorporate continuity rules identifying the types of legal and administrative signals that result in the birth, death, or continuation of an enterprise (or establishment). These rules determine under what circumstances an enterprise is deemed to be born, under what circumstances it is deemed to have died and possibly replaced by the birth of a new enterprise, and under what circumstance it is deemed to continue but in some new form, or under new ownership.

In order to be practically applicable, the rules have to take into account the ways in which changes can be detected.

The rules have to cover every possible type of birth, death, or continuation of units that can occur, taking into account the signal source and the maintenance group to which the enterprise belongs.

#### *Attribute change stability rules*

Changes to the enterprise attributes used for sampling must be subjected to resistance rules, which ensure the change is permanent before an update is made.

These rules must be defined taking into account how the changes can be detected, the type of change (permanent, temporary, seasonal), and the impact on published statistics.

The rules have to cover every possible type of change that can occur, taking into account the attributes involved. They are likely to include resistance rules that inhibit short-term changes.

#### *Dealing with conflicting information*

In order to deal with conflicting information from different sources, the sources must be prioritized according to their reliability.

## 7.5 SBR maintenance: quality management aspects

The goal of quality management of the SBR is to meet the requirements of the users. Quality can be defined through the dimensions of relevance, accuracy, timeliness and punctuality, accessibility and clarity, coherence and comparability. There is a potential conflict between user needs (compiling statistics) and the view that the content of the SBR should reflect reality faithfully (complete and up to date coverage and content). Statistical business registers should be always seen in relation to their use in the statistical production process. The definition of quality targets and measures must be subjected to a cost-benefit analysis.

The management of stakeholder requirements includes such aspects as advance knowledge of the surveys that will be drawn from the SBR, the monitoring of which units are surveyed and for what statistics, a helpdesk to queries related to the sampling frame and the statistical units and the management of feedback from the stakeholders. These aspects can be supported by tools such as the identification of units, the access to the business register, the coordination of surveys and information on significant changes in the population

Data source management requires service level agreements with the administrative sources (roles and responsibilities, data specifications, delivery formats, dates, quality), knowledge of the methods and processes followed by administrative data suppliers, data validation rules and rules for the integration of multiple data sources (priority settings for conflicting information)

Metadata must include the source of data items (statistical survey, administrative source, etc.) and time stamps (reference date, date of last update, etc.). The source of data items and dates should be available at least for the main stratification variables (employment, turnover and economic activity classification).

## 7.6 Treatment of errors

The chapter 18 of the EU Business Register Recommendation Manual gives an extended definition of errors and how to handle this.

It is impossible to avoid errors and incorrect information in a business register and it may not even be desirable, for reasons of consistency, for the register to be perfectly accurate at a particular point in time.

Changes to information in business registers can either be a reflection of real world events, in which case they are regarded as updates, or they may amend information that was previously wrong, in which case they are regarded as corrections.

### *Definition of errors*

The Eurostat manual on Business Register define an error as "a difference in the information presented in the register and the information as it should be, according to a chosen image of the real world produced and maintained by an accepted instrument and documented procedures" (See chapter 8 Quality of SBR)

### *The handling of errors for different users*

The handling of errors should take into consideration the different types of surveys based on the register. The solutions acceptable for structural surveys may be different to those appropriate for short-term surveys. Where different surveys demand different solutions for the handling of errors, it

may be useful to record details of errors including dates of detection, occurrence and correction, sources and types. This enables register users to access both corrected and uncorrected data according to their needs.

One approach is to set up a specific database to record these details. The structure and functionality of such a database should depend on the procedures agreed for handling errors. This database should be closely linked to the business register or even an integral part of that register. The following variables should be recorded:

- The original value (the wrong one)
- The new value (the correct one)
- The date of detection
- The date of occurrence
- The date of correction
- The source of the corrected value
- The mode of correction (interactive or batch; in the case of interactive correction, it may be wise to record the corrector as well).

It may also be useful to record a date of confirmation for variables that may be mistaken for errors because they look unusual or out of date, but have been investigated and confirmed as correct.

If possible, the type of error should also be recorded for further analysis. The inclusion of all this information in the database makes it possible to recreate the situation in the register at any given moment, allowing users to ignore corrections if necessary. The inclusion of additional variables in the business register also has clear disadvantages, more variables make register maintenance more complex and mean more opportunities to make errors, but the advantages of at least partly implementing this solution should more than outweigh the disadvantages.

Special provisions for the handling of errors in large and important units may need to be negotiated with users. Errors in important units may have a considerable impact on certain types of statistics. If a special procedure is implemented for these units, it should be fully documented in terms of the units to which it applies. Special provisions can be applied through a dedicated profiling team or by appointing an authority (e.g. the head of the business register), who has the final say in these corrections.

#### *The time dimension*

If errors are detected, it is important to know the exact moment at which the error occurred. If this date and the date of correction are recorded, it is possible to construct populations of units for certain points in time that may be more accurate than the situation in the register at that point in time. The date of correction, if a correction is carried out, is likely to differ from the moment of occurrence. In practice, the date of occurrence is often not known.

A very important question is "How far back in time should corrections be applied?". The answer depends on the correction strategy not only of the register but also of the statistics based on the register. If, for example, there is a procedure of revision in which all statistics are revised after a period of five years, then correction should be carried out to the moment of the previous revision. If there is only a revision procedure for national accounts and not the individual statistics on which they are based, the period for which corrections should be carried out may depend on the frequency of the statistics based on the register.

Another important consideration is the moment at which corrections should be made. Corrections should be applied according to a fixed and documented procedure wherever possible. Corrections of errors in identification characteristics, such as names, addresses, telephone numbers, etc., should normally be carried out immediately. If not, they may hinder data collection and possibly also frustrate respondents.

The correction of errors in economic/stratification characteristics, such as economic activity and size, is a different matter. If these corrections are made the moment the errors are detected, they will affect consistency between statistics with different periodicities. To avoid this, the corrections should be stored until an agreed moment of correction, for example the first week of January of each year. However, if the register records information which allows survey statisticians to ignore corrections where necessary, there is no need to delay the application of corrections

#### *Sources for the correction of errors*

The source from which a correction originates is an important factor in deciding whether or how to correct the error. Care must be taken to ensure that the register remains a known reflection of reality and that corrections do not lead to unknown distortions of that image. If the source is a statistical survey based on the register, corrections may lead to bias depending on the particular sampling scheme used. This is because the corrections would only apply to a part of the register, whereas for other parts, where units were not included in the sample, the quality would remain as before, resulting in different quality levels in different parts of the register. (e.g. the Prodcom survey, which only covers production units). Such a survey will help to identify units that wrongly fall within the scope of the survey, thus allowing these errors to be corrected, but it will not identify units wrongly out of scope. Applying corrections on this basis will mean that the quality of in-scope units will generally be higher than that of out of scope units. There would also be a distortion in the distribution of units in the register. The population of in-scope units will be lower than it should be, whereas that of out of scope units would be correspondingly higher.

Corrections from sources that are comprehensive in nature can be processed without problems, as it may be assumed that they are randomly distributed over the units in the register. Corrections from all other sources will introduce a certain level of bias, and it is therefore necessary to determine whether this is acceptable and preferable to users.

In summary, there are three steps to the handling of errors:

- Decide whether errors have occurred
- Decide whether they should be corrected
- Decide how and when to correct them

Errors will occur in all statistical business registers. To reach a level of quality acceptable to users, it is necessary to develop a systematic policy for the handling of errors. The following recommendations aim to help the formulation of such a policy:

- It is helpful to make an inventory of register uses and the consequences of different types of errors for different users.
- The business register should be structured and maintained in such a way that the correction of errors has a minimal impact on statistical surveys.
- Register inputs, processes and outputs should be systematically monitored to detect potential errors.

- Register processes should be fully documented so that the correct handling of all variables is clear to all concerned. This is necessary to detect errors and avoid discussions on the quality of individual records.
- The policies regarding the handling of errors in the register should also be fully documented, and audited periodically to make sure they are still appropriate.
- Responsibilities regarding the handling of errors should be clear and documented. It is advisable to appoint an authority (e.g. the head of the business register) who has the final say in difficult cases.
- The handling of corrected register values in statistics based on the register should be fully coordinated and documented.
- The different types of errors detected should be analysed periodically to monitor changes in the pattern of errors over time and thus to inform on the development of policies for handling errors.
- Recording the history of errors will facilitate the handling of errors in complex situations.
- If statistical business registers are used for administrative or commercial purposes, it is advisable to take legal precautions against damage claims in the event of errors