

Regional Training on Producing Register-based Population  
Statistics in Developing Countries  
23 September – 31 October 2013

**e-learning module: *Basic information and statistical background***  
23 – 27 September 2013

Lesson 2

---

**General preconditions for using administrative  
sources for producing population statistics**

---

**References and complementary readings:**

1. United Nations (2011): [Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices](#)
  2. United Nations (2007): [Register-based Statistics in the Nordic countries](#)
  3. Wallgren, A. and Wallgren, B. (2007), *Register-based Statistics: Administrative Data for Statistical Purposes*, John Wiley & Sons, Ltd, Chichester, UK.
-

---

## Contents

1. Requirements to data from administrative sources
  2. Frameworks for access to administrative sources
- 

### **1. Requirements to data from administrative sources**

This lesson describes some demands that must be met if official statistics are to be successfully based on administrative registers.

#### *Contents*

Administrative registers must contain data covering the most important subject areas in a statistical system for elucidating patterns and trends in society. An important precondition for statistics based on registers giving comprehensive coverage, is that *the data contained in administrative registers should be extensive and should cover many variables relating to the relevant units*. Gaps either necessitate the supplementary collection of information using traditional methods or limit the content of the statistics.

#### *Units and identifiers*

Three central units are essential to the structuring of the statistics: persons, enterprises/establishments and dwellings. There are of course many other units for which it is important to have records. Some of these units are events, which relate to persons, for example demographic events (births, deaths, marriages etc.).

For statistical use, it is important that the units are well defined. This is not a problem for "natural" units such as persons or motor vehicles. Also, the concept of a home or dwelling is fairly easy to deal with, although there are some borderline cases in which the existence or specification of a home can only be determined on the basis of detailed rules. Some units manifest themselves as combinations of natural units, e.g. families and households.

It is more problematic when other types of units are considered, and here business units are a good example. A firm or a workplace is an abstract concept. Its existence can only be established and its nature defined in accordance with very precise rules. For instance, from administrative data it may be difficult to determine whether a new enterprise has been established or if an existing one has undergone some major changes.

Registers should, ideally, contain information on all units in the country of the types specified, i.e. *they must be comprehensive*. Incomplete coverage will often be systematic, for example, if certain municipalities are not covered, the statistics produced on such a basis will immediately be affected by biases that it may be difficult to compensate for.

Identifiers play a very considerable role both in the maintenance of administrative registers and in their statistical use, particularly in linking information from various sources. 'Identifying' codes should ideally not be changed for the period a unit exists. *This is best achieved by using codes containing no information*. When a personal identification number contains the date of birth and sex, there may be changes during the lifetime. The most

---

frequent reason is that the date of birth is being corrected in the Central Population Register (CPR), mainly for emigrants. Problems may also arise when a person arriving in the country is initially given a temporary identification number and subsequently a permanent personal identification number. In such cases, the NSO has to establish links between identification codes that refer to the same person. One way of solving this problem is to replace the administrative identification number by a statistical identifier in statistical registers.

### *Time references*

The time dimension plays a very important role in statistics, revealing patterns and trends in society, and in all areas it is necessary to be able to make comparisons over time. It is therefore vital for statistical usability that reliable time references are contained in registers.

The most important is the dates of changes or events. Among the main events of interest are the "birth" and "death" of units, but it is also important to date other changes affecting units. What we are concerned with here is the real point in time at which an event took place bringing about a change in a data item, for instance the date of a removal or the date of a change of industry for a business enterprise.

In practice it is often difficult to pinpoint the real date of an event. The industry within which an enterprise operates is a variable that may change gradually, as the activities of the enterprise shift their emphasis, and it is perhaps not possible to determine exactly when a new activity becomes dominant.

However, even for variables that in principle can be observed at any given time, and in which a change can therefore be dated with reasonable accuracy, problems arise in practice in the recording process. This may be due to the fact that the date of an event is not necessarily important to the administration using the data: all it needs is to know the most up-to-date value at any given time. In other cases problems may be due to the fact that it is not possible or practicable to record events in real time. It may only be possible to ascertain data changes from responses to enquiries, for example once a year in conjunction with a tax return.

In addition to dates of events there is a need for registration dates, i.e. an indication of when the data value in question was entered in the register. The ideal situation therefore is that any item of information in the administrative register should be accompanied by two dates. Registers in reality often deviate significantly from this ideal.

If it is not possible to record dates with accuracy, approximations of course remain an option and are much better than nothing at all.

### *Data on events*

In some cases there is a requirement that it should be possible to view the data changes recorded in the register as a model of events, which are described in the statistics. This applies for example to population statistics, in which it should be possible to derive reports of events in the population (removals, marriages etc.) from data changes in the Central Population Register (CPR). For the purposes of migration statistics for instance, it must be possible to distinguish real events (removals) in the register from corrections and data changes of a more technical nature.

---

Actual corrections of wrongly reported data however pose a problem for statistics, even when it is possible to distinguish them from real events. Corrections have to be regarded in many cases as modifications of events reported earlier: an address correction may be a correction to a report on a removal. A correction may also cancel an event reported earlier, which occurs in the CPR system in particular when a person, having notified the population register that he is to emigrate, changes his mind and remains in the country.

### *Stability*

An important characteristic of statistics is to describe a process over time, i.e. to show how a particular magnitude develops from month to month and from year to year. It is of great importance, therefore, that concepts in the administrative registers remain constant over the longer term. Otherwise major problems can arise in securing comparable figures from one period to the next. In some cases, but far from all, it is possible to adjust for changes with greater or lesser precision. Paradoxically, a problem may arise even if the registers become more reliable, for even then we get data discontinuity. Major problems may be posed for statistics when legislative or regulatory changes result in alterations to the data content of administrative registers. On the one hand, it may be difficult or impossible to assess the long-term trend in a particular magnitude if different definitions are used in the base material. On the other hand, problems may arise in deciding what changes in data values are to be viewed as reflecting actual events and what changes merely represent new concepts or definitions.

The statistical consequences depend on what type of statistics is involved. If it is in fact statistics for monitoring of legislation, the function of which is to show how the administration of a law affects ordinary people, the statistics of course merely have to go along with and adopt the concepts of the new legislation. General statistics, on the other hand, seek to elucidate certain concepts that are not defined in legislation, e.g. unemployment, a concept whose definition is to be found in an international convention. If the statistics relate to the payment of social benefits, and the rules for these are changed, it may be difficult or impossible to compensate for the change in the statistics. An attempt must at least be made to estimate the significance of the changes, so that time series can to some extent be chained together.

There are also instances of administrative registers changing without this being due to any changes in the basic rules. Typically these are changes introduced for technical reasons or to achieve rationalization gains. A problem may arise, for example, if it is decided that a particular item of data is no longer necessary for administrative purposes.

Clearly the changes described are serious for the statistics if they are undertaken without sufficient consideration for its requirements. It is therefore important that statisticians are involved in the preparation of the changes so that at least their consequences can be assessed before they are introduced.

### *Quality*

The quality requirements imposed by statistical use coincide to some extent with the requirements that must also be met in serving the primary purpose of the registers; the

---

information must be reliable and recorded with sufficient precision. In addition, data must be relevant to the statistics. Quality aspects are discussed in lesson 1.

### *Cooperation with register-keepers*

In producing statistics based on administrative data it is not possible to exercise the same control over the content of basic data as in the production of questionnaire-based statistics. We cannot be sure that the registers cover the units of relevance with the same degree of precision or that data are defined in accordance with the needs of users.

It is of course desirable for statisticians to exert a certain influence on data content, but it must not be forgotten that *registers are kept for quite specific administrative purposes and that the task of the register keepers is to serve those purposes in the best way possible. They cannot therefore pay too much attention to demands that "only" serve statistical purposes.* For this reason, statisticians should not expect to make very extensive demands for additional data, different definitions and the like, and to have those demands met. Register keepers must pay strict attention to efficiency in their own operations and use of resources. Statisticians must therefore be modest and only put forward major requests when an adjustment to a register can be expected to yield very substantial benefits to society. Very few instances of extra data being collected by way of registers exclusively for the purpose of statistics are known in the, for instance, Nordic countries. A national statistics act is a useful instrument for good cooperation with register owners.

## **2. Frameworks for access to administrative sources**

The access to data from administrative sources is one of the key barriers to the wider use of such data for statistical purposes. This section describes the various frameworks needed to facilitate access to administrative sources, drawing on examples and experiences from several countries. These frameworks typically have several dimensions; legal, policy, organisational and technical, each of which is described below.

### *Legal frameworks*

*Legal frameworks* are normally constructed at the national level, and are specific to national sources and circumstances. In some cases, however, there may also be relevant legislation at either the sub-national (e.g. state) level, or the international level. An example of the latter is the statistical legislation of the European Union, which is binding on Member States.

Most NSOs have legal texts defining their roles and responsibilities, typically in the form of a statistics act. In many countries, these legal texts include specific provisions for the access to administrative data (examples are Ireland and Norway).

The legislative process can take time, and statistics may often be seen as a relatively low priority by legislators, so a sustained period of lobbying and highlighting the benefits of using administrative data may be necessary. Given all the efforts that are usually needed to introduce or revise statistical legislation, it is therefore necessary to make the most of the opportunity.

Even whilst legislation remains a barrier, it does not necessarily prevent any use of administrative data.

---

### *Policy framework*

Many countries have general policies on data sharing within government, which will influence the right of access to administrative data for statistical purposes. However, it is often easier to change policies than to change laws, and policy tends to evolve over time. It is therefore important that NSOs participate fully in policy development, and take an active part in any discussions within government that might lead to policy changes. In this way, any changes should be formulated in a way that gives the maximum possible benefit to the statistical system.

Policy frameworks also encompass voluntary codes of practice, the most important of which, for statistical purposes, is the United Nations “Fundamental Principles of Official Statistics”. Principle 5 concerns cost-effectiveness, and suggests the use of data from administrative sources in this context.

### *Organisational framework*

Once the legal and policy frameworks are in place to permit the use of administrative data, it is necessary to consider the organisational arrangements to facilitate data flows. Typically this takes the form of a written agreement. This may be a contract, particularly if a private sector organisation is involved, but, if the agreement is between government departments or agencies, it is more likely to be a “service level agreement”, “protocol” or “concordat”. The difference is that contracts tend to be legally binding, whereas other forms of agreement are not.

### *Technical frameworks*

*Technical frameworks* refer to the mechanisms by which data are transferred, as well as any relevant data or metadata standards. Data transfer mechanisms can take any form from paper records sent by post to real-time updates via a secure electronic link. The mechanism used has to take into account the technical possibilities open to both the sending and the receiving organisation, so is often a compromise reflecting a sub-optimal solution for at least one of these organisations.

There are a number of international standards for data and metadata transmission, including XML, SDMX and DDI, to name but a few. Some countries also have national versions, particularly for data transfers within government. It is therefore important to agree which standards are to be used.

***It is essential to have a legal framework in place to permit the use of administrative data for statistical purposes. The other frameworks described above are not essential, but are very useful for assuring a smooth flow of data, and minimising any problems or misunderstandings between the data supplier and the statistical organisation.***