Kåre Vassenden

8. The data basis for Statistics Norway's migration-related statistics

Production of comprehensive and reliable statistics in the field of international migration and immigrants requires the existence of relevant data, that the data are available to the statistical office, and that it is legally and technically possible to process the data and make them suitable for the production of statistics.

8.1. The data collection

In general, Statistics Norway obtains data for the production of person statistics from

- nationwide administrative registers

- institutions or agencies covering limited geographical areas

- direct collection from the informants via interviews, postal surveys and complete counts.

The data collection from the registers and institutions is based on Statistics Norway's right pursuant to the Statistics Act and the use of such data sources for the production of statistics.

The administrative register data form the basis for drawing survey samples, and information from the register is also linked to supplement the survey data. Thus, access to good register data is a basic requirement for all Norwegian migration-related statistics, even those that are based on surveys.

In addition to the data retrieved externally, Statistics Norway has considerable amounts of data from earlier data collections. These data also represent an important source for the production of statistics.

Statistics referring to immigrants are mainly based on counting people. In other cases the unit is events (for instance migration events) or other circumstances related to persons. Almost all accessible data for such statistics are at individual level, i.e. the data contain information on each counting unit.

8.2. The Norwegian register is integrated and comprehensive

What characterises Norway and some other countries (primarily the Nordic countries) with regard to administrative register sources is that they cover many aspects of society, that the quality of the registers is relatively high, and that there is contact between registers that enables a certain degree of coordination. The principle is that only one register is responsible for a certain kind of information, and that the data are then exchanged between the registers that need the data.

The Population Register functions as a hub for all person registers. Both the Aliens Register in the Norwegian Directorate of Immigration and the Adoption Register supply data to, and are users of the Population Register. The system of population registration is a part of the Tax Administration, which is also responsible for other registers related to taxes. Last but not least, the Norwegian Welfare and Labour Administration (NAV) has several large registers related to national insurance and the labour market.

8.3. The population register has a key role

The Population Register plays a key role with regard to person data in Norway. As with similar registers in other countries, the Norwegian Population Register assigns ID numbers, and registers central demographic and legal person information.

Only in a few countries, however, is the concept of residence as strong as in Norway. Not only does it entail citizens having to give notification every time they move, it also means that the Population Register constitutes a public authority, which through legislature decides who is resident in Norway (and where in Norway they are registered as resident). All other public agencies are obliged to base their case handling on this definition of the resident population.

The body of rules on where persons shall be registered as resident is based on where they take their regular night-rest, but there are several exceptions to this main rule. The result is that the registration of residence does not reflect the actual situation as much as it would with another body of rules.

The biggest challenges, however, are linked to moves that are not notified to the

Population Register; this particularly applies to emigration events. Cases are constantly being discovered where someone has left Norway for such a long period that they should have been registered as emigrated.

8.4. The most important ID number series

Most of the individual data that Statistics Norway receives uses the Personal Identification Number (PIN) as identification. The so-called D-number is also used in certain cases, in addition to other number series that are used in relation to data from the Immigrant Administration.

PINs are assigned by the Population Registry and registered in the Central Population Register (CPR) database. Everyone born in Norway receives a PIN, even those who are born on Norwegian soil without resident parents. The same applies to immigrants who meet the conditions for being registered as resident in Norway. In addition, some smaller categories are given a PIN even if they have never resided in Norway (e.g. children of Norwegian citizens abroad).

At the beginning of 2010 there were almost 7.6 million PINs in the CPR, covering 7.4 million people.

The reason for the difference of slightly more than 200 000, is that some people have had more than one PIN during their life. Most of the changes of PIN have taken place in order to correct the date of birth. In particular, such corrections happened in the first years after the establishment of the register and the PIN in 1964. Every year a few dozen cases are discovered where two valid PINs actually are one and the same person.

The CPR contained 4.7 million residents at the beginning of 2010, more than 2 million deceased persons and 470 000 people who had emigrated since 1964.

The D-number is assigned to persons who do not qualify for a PIN, but nevertheless have economic relations with the Norwegian authorities. The Tax Administration is also responsible for this number, and the D-number register is part of the CPR database. More than 1.3 million D-numbers have been assigned through the years. The D-number has the same format as the PIN, with the exception of the date of birth (the two first digits), which is increased by 40. If someone who only has a D-number immigrates to Norway (i.e. is registered in the Population Register as resident), a link from the D-number to the PIN received at the time of immigration is established. In many ways, the PIN and the D-number belong to one and the same ID number series. Otherwise, very little information is connected to the D-numbers.

In the Norwegian Directorate of Immigration, a DUF number is assigned to almost everyone who is registered in the Aliens Register. For those with both a DUF number and a PIN (or D-number), there are arrangements in place to ensure that this connection is stored in both the UDB and CPR. The coverage of old registrations, however, is not as good as could be hoped.

8.5. The data sources of the different subject areas

The different subject areas in Statistics Norway use, in most cases, one or a few main sources, in addition to several supplementary and complementary sources.

The CPR is the dominating source for the population (demographic) statistics, but data from the Aliens Register, Adoption Register and own historical data are also utilised. For instance, data on country of birth from the Population and Housing Census 1970 are used to supplement the information from the Population Register. In order to construct the variable «reason for immigration», data from the Aliens Register are used, in addition to own data.

The education statistics are partly based on administrative systems, partly on data directly from the educational institutions and from bodies such as the Norwegian State Educational Loan Fund and NAV. The latter agency, with its Employees Register, is the most important supplier to the labour market statistics, which also utilise the Tax Return Register and the Register of End-of-the-Year Certificates from the Tax Administration. The income statistics also use other registers from the Tax Administration, such as the Tax database of personal tax payers, and NAV's payment registers.

The major register sources are valuable for the production of statistics, but they are rarely without faults or weaknesses, even in the eyes of the administrative agencies themselves. Some of the weaknesses that the producers of statistics point out are linked to the use of data that have been adopted for the purpose of handling administrative cases. In public administration, no information is recorded unless it is correct within a degree of certainty, because the data may have consequences for individuals at a later stage. In addition, information that is not necessary for the immediate case handling may not be registered. Statistical considerations may easily be given less priority if extra measures are not imposed.

The coverage of registers is rather high, but there are some holes that create

problems for the producers of statistics. One of the biggest problems is the lack of a complete system for collecting information on the education that people bring with them from abroad.

8.6. Processing the data at Statistics Norway

The data basis for immigrant statistics is normally only one part of the data basis used for the production of the statistics for the total population, but some data sources are particular to persons with an immigrant background.

As mentioned above, the administrative data that Statistics Norway receives have properties specific to administrative case processing. When compiling statistical data, considerations other than the administrative ones should and can be taken. In the production of statistics for instance, a low number of non-stated cases are given more weight than the reliability of the data for each single person. The statistician may refer to the law of big numbers, which means that small and random errors are counterbalanced, and for that reason can be ignored.

When data are received by Statistics Norway they are subject to extensive processing before they can be utilised in the production of statistics and/or as input for other data development projects. Data are checked and repaired or improved in different ways during the processing stage. New variables that are more suitable for the statistics are often produced, and information from previous editions or other sources are often matched up.

The quality can be improved when utilising data from several sources. Matching is also necessary when data from different fields are merged. When combining data in new ways, new knowledge can emerge.

Extensive matching between different sources can often lead to difficulties in selecting only one specific source for a certain processed variable or complete data set.

The PIN has been a key in the production and development of social statistics in Norway. The existence of this ID number series in practically all relevant input registers enables the simple and secure forming of links.

The result of the processing is often important data collections that are given names such as the population statistics system, register-based employment statistics, the income registers or similar. These are the direct sources of the immigrant-related statistics.

8.7. A system of data files

In Statistics Norway, the base variables are produced in one place and then distributed to the subject areas that need them in their production of statistics. There the variables are linked by means of the PIN. For example, who are classified as immigrants is decided one place, the level of education another place, and income a third.

The system is integrated in the sense that it is technically easy to perform matching when necessary. It does not mean, however, that there is no control of what is matched across the subject areas.

This exchange of data between subject areas and the preparation of data for research projects mean that, in principle, there are three objectives with data production in Statistics Norway: the data shall be suitable for the direct production of statistics, as input to other data development processes and for internal and external research projects and other commissioned projects.

All data at Statistics Norway eventually becomes different kinds of statistics, but prior to that stage data are needed as individual level data and with an ID number series as a linking key. Each of these data files has many uses.

From a legal perspective, data from administrative primary registers and separate collections become «statistical registers» as soon as they are received by Statistics Norway. This means that these data shall only be used for the production of statistical data and statistics.

Technically, registers and variables do not become statistical until they have been through a preparation process that converts them into something other than the original, administrative data.

8.8. The classification of persons by immigrant background

In Statistics Norway, what is classified as immigration and who has an immigrant background is technically defined in the population statistics system. Central immigrant statistics variables are citizenship, change of citizenship, country of birth, immigrant category and country background, reason for immigration and first date of immigration. New variables are constructed from these, such as the length of residence in Norway and age at immigration.

From the «core» of migration statistics, represented by files with country of birth and other population data, individual level data are forwarded to subject areas such as education, population census, health, national insurance, crime, labour market, income, interview and so on, for linking to the data files of these subject areas.

A prerequisite for this model is that all relevant subject areas have good specific files and that it is easy to match up variables from other subject areas. When that is the case, definitions of immigrants and information about them can be produced in one place and then matched to all the files that are to be used for statistics by immigrant background.

In this way, all person statistics can be distributed by various immigration-relevant variables, which is not always possible in countries with different systems of statistics. In addition, one result is that precisely the same persons with immigrant background occur in the various statistics, which are thereby comparable across the subject areas.

8.9. Personal protection

Statistics Norway has responsibility for many statistical files containing, in part, sensitive information on the total population or parts thereof. The use of registers at Statistics Norway is governed by the Statistics Act and the Personal Data Act. Statistics Norway has its own personal protection ombudsman who ensures that the handling of personal information is performed correctly, and that routines aimed at securing this information on individuals or small groups are not compromised by the publishing of statistics.