Annual "census" migration data in a developed register-based system – the case of Slovenia

Barica RAZPOTNIK

Senior Advisor, Statistical Office of the Republic of Slovenia, Slovenia

Danilo DOLENC

Under-secretary, Head of Section, Statistical Office of the Republic of Slovenia, Slovenia

Abstract. The Eurostat document "A vision for the census statistics after 2021" is a starting point for elaborating international migration issues from producer's point of view (quality and relevance of administrative data used, integrated statistical production and re-use of statistical data for migration statistics, appropriateness of definitions used in migration statistics) and user's needs (more frequent - annual data; more topics with stress on socio-economic data and population with immigrant background; shorter time for data production; availability of data on detailed regional level). The good solutions and problematic issues will be also exposed.

1.   Migration data in a register-based census

The main (and in some surveys only) data source used in demographic statistics is the Central Population Register (hereinafter CPR) from which we receive regular quarterly transmissions of data on all demographic events (flow data) as well as stock data on persons who have ever been assigned a PIN regardless of their status of the residence in Slovenia. CPR has been in fact established by the Statistical Office of the Republic of Slovenia (hereinafter SURS) in the 1980s but transferred to relevant ministry (Ministry of the Interior) after the adoption of the National Statistics Act in 1995 following the general recommendation that administrative function of a register must be separate from the statistical one.

Migration statistics in Slovenia have been fully register-based since 2008. Since then SURS has been receiving individual data on all persons registering or deregistering residence (prior to that data on immigration of foreigners have been received in aggregate form while data on emigration of foreigners had to be estimated). In combination with individual register-based data on births and deaths, this results in completely balanced annual stock data where all individual entries and exits are accounted for. It was just a coincidence that at the same time Regulation (EC) No 862/2007[1]

---

[1] Regulation (EC) No 862/2007 of the European Parliament and of the Council of 11 July 2007 on Community statistics on migration and international protection and repealing Council Regulation (EEC) No 311/76 on the compilation of statistics on foreign workers (CELEX: 32007R0862)

was adopted which determined the basic definitions and concepts of usual residence, immigration and emigration that were fully applied in Slovenia since 2008 using intended stay in the country as the main principle for determine belonging to usually resident population. Having migration statistics data in mind the short term and long term immigrants could be distinguished based on intention of stay in the country and only long term immigrants are included in the population stock. In order to maintain the coherence with other population statistics, the census definition in 2011 Register-based Census was harmonized with regular (quarterly) population statistics for the first time. In fact the pre-defined and statistically processed population data from only one administrative source (CPR) prepared according to the EU Regulation on migration from 2007 is the basic input database (usual resident population) in the register-based census process.

During the 2011 Register-based Census, several migration-related variables were derived. Some of them were obligatory as per Regulation (EC) no 763/2008[2], namely:

- place of birth (meaning the place of usual residence of the mother at the time of the birth, or, if not available, the place in which the birth took place);
- country of citizenship;
- ever resided abroad and year of arrival (calendar year in which a person most recently established usual residence in the country);
- place of usual residence one year prior to the reference date;
- previous place of usual residence and date of arrival in the current place.

Additionally, UNECE Recommendations listed further international migration-related topics of great interest. The following were collected / derived and are available to users from our census database:

- the year of first immigration to the country for foreign-born population;
- immigrant background for all usual residents (so called ''first / second generation of immigrants'');
- country of birth of parents (for population with established links between children and parents).

The statistical process for the ''census'' migration data has been established in the scope of 2011 Register-based Census combining data from different sources:

- 2002 Population Census (data for the last time collected in the field using the place of usual residence of the mother at the time of the birth as a place of birth);
- Annual statistical survey on migration (data from 2002 to 2010) - de facto country of birth generally considered as place of birth for immigrants;

---

[2] Regulation (EC) no 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses (CELEX: 32008R0763)

- Annual statistical survey on birth (data from 2002 to 2010) - if the link between mother and child exists the place of mother at the time of birth considered as place of birth; if not the first address of a child in the CPR;
- Quarterly statistical data on population, as of 1 January 2010;
- CPR (citizenship, links between children and parents).

Due to methodological differences for specific migration topics, the statistical process has been adapted for each topic separately. The basic methodological principle is the hierarchy of the sources. In case of availability of data for the same person in several sources, the priority is given to the source indicated with higher priority, or in case of more than one record for the same person in the same source (valid only for statistical survey on migration), the priority is given to the methodologically adequate data.

2. Annual "census" migration data

User needs for more frequent updates on these topics and a more automated process along with better availability of quality individual-level data sources have led to regular annual production of data on topics previously only collected with the census. As mentioned, data derived in 2011 Register-based Census serve as the basis for the derivation of these variables, but additional data sources are used as well for updates, if needed.

Some migration characteristics are not changeable (such as place of birth) but data could be changed anyway (in case of territorial separation or territorial fusion). On the other hand, every new record (entry to usual resident population) must be assigned a place of birth and year of first immigration (if applicable). For infants, data mostly come from the annual live births database. For immigrants, data are derived from previous reference dates (if available) and from the annual migration flow databases. These annual updates are available approximately eight months after the reference date.

Deriving data on immigrant background and the country of birth of parents (which is now also considered a core social variable and should be collected in all social surveys according to the new IESS Regulation) is a somewhat more complex procedure as a person's links to parents (PINs) are needed. For older persons and foreign-born foreign citizens these links are often missing, but the coverage of these data is improving rapidly. In 2011 Registered-based Census, the share of available PINs of mother was 73% and PINs of father 66%. From the point of view of current migration movements, population aged less than 50 years is more relevant. For population aged 0-49 years which represents 60% of the whole population, the coverage of PINs of mother has reached 91.5% in 2017 and for PINs of father 89.5%. In addition, the parents, even if known, may

never have had their place of birth derived in our population stock databases, and information on them in the CPR, which serves as proxy in such cases, may be of lower quality.

Additional migration-related variables are also derived annually. Since 2011 we have been preparing register-based annual data on current activity status using methodology which is very comparable with Labour Force Survey (hereinafter LFS) definition of main labour status (90% of persons had the same labour force status in both surveys) and on educational attainment. Before 2011 census, these important topics were frequently only available from the LFS where immigrant population is generally under-represented. The share of foreigners (most of them are recent immigrants) aged 15-64 years (working population) in LFS in 2015 was 3.6% while the share of the same population in 2015 Register-based Census was 6.1% (almost twice higher). In population database the coverage is complete, and every usual resident of Slovenia and every international migrant or person with immigrant background aged 15 or more must have information on the level of education and current activity status.

Data on socio-economic characteristics are derived from eleven different sources, all available at individual level (statistical registers, statistical surveys based on administrative data, databases of institutions with relevant data etc.). These annual updates are published 11 months after the reference date/end of reference year. Once data are derived for the whole population the same data are re-used for other demographic and social sample surveys without asking the respondents questions on topics that are already available from administrative and statistical sources. This leads to higher consistency of data, to better harmonization and standardization of the statistical processes and allows the macro linkage of data from different statistical surveys.

The current activity status data derived from administrative and statistical sources in register-based population statistics = census statistics are also used for measuring over-registration in basic data source (CPR). The imputation rates for current activity status (no data on individual person in any of the sources, only data from the CPR are available, so there is a high probability that those persons live abroad) are decreasing year by year (1.50% in 2011, 1.17% in 2015) and are sign of improved quality of the CPR data. Beside that a special survey on potential non-registered emigration was carried out in the scope of grant ''Usual Residence Population Definition: Feasibility Studies'' funding by Eurostat. Results of the survey confirm our presumptions (low response rate, over-registration in the Central Population Register due to undocumented emigration is bigger than under-coverage of returned emigrants). The main lesson learned from the survey is that population data are of adequate quality for statistics (over-registration around 0.5% or around 10,000 persons due to non-registered emigration) which is much less as the coverage error was in

the last traditional 2002 Census (around 2% of under-coverage and 1% of over-coverage due to double counting).

3. Key conditions for the production of annual census migration data in a register-based environment

As with all register-based statistics, an established PIN is the basis. Ideally it is a number that every person in the country knows and uses regularly in different situations (enrolling in education, applying for social assistance, registering residence etc.). A person without a PIN does not exist in a statistical term. This identifier must be used consistently in all data sources, meaning that the same person appears in different sources with the same identifier. This is the key for linking data from different sources. An alternative option can be identifiers that are mutually "translatable", for example tax ID that can be uniquely linked to PIN.

Having an established system of data transmissions from other institutions is equally as important. Formal agreements with data providers ensure consistent and timely transmissions but are not sufficient. SURS focused most of its efforts on the preparation phase of input databases to inform data providers of the importance of sharing responsibility for accurate results.

For census migration data production, the existence of connections between parents and children is also important, i.e. PINs of parents for every person in the CPR. This is the key when determining the country of birth of a person's parents. It is also useful for other similar derivations, such as educational attainment of parents. In fact every characteristic of parent (if they belong to usual resident population) could be linked to their children.

Even in cases where PINs of parents are not available, connections between persons can be derived from relations within a household. This is only possible if household data are derived as well, but even data on relations from past derivations of data on households could be used. In the post-2021-census perspective derivations of household data are likely to become a more frequent procedure, possibly annual.

Individual data from the last classic field census are also very useful, especially for the first execution of register-based census. Since the values of core census migration variables generally do not change (i.e. a person's country of birth does not change, nor does the year of their first immigration), they can be used for later iterations of the process.

4.   Conclusion

The register-based approach in the production of migration stock data keeps all essential census features (universality, individual enumeration, simultaneity, availability of small-area data), with an important improvement in periodicity turning to annual updates of all topics from EU Census Regulation and also additional (mostly derived) topics which gives our users quality and up-to-date harmonised, reliable and relevant statistical data on different population groups relevant to international migration (foreign-born, foreigners, descendant of foreign-born, ever international migrants).

The developed system of migration and other census topics already completely satisfies the EU vision for the post-2021 census, with the following key features (Eurostat, 2016):

- Annually updated data;
- Annual data to be available for detailed geographical levels (NUTS3, LAU2), with annual population counts (persons by place of usual residence) available also for a geo-referenced grid (in our case the smallest grid used for dissemination of data is 100x100 m2);
- Annual data to be available within 12 months of the reference date;
- A more detailed and comprehensive (relative to the annual data) decennial (2031) data collection to be retained.