

Guidance on Data Integration for Measuring Migration



UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

Guidance on Data Integration for Measuring Migration



UNITED NATIONS

New York and Geneva, 2019

Note

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

ECE/CES/STAT/2018/6

UNITED NATIONS PUBLICATION

Sales No.: E.19.II.E.7 ISBN: 978-92-1-117185-3 e-ISBN: 978-92-1-047575-4
--

Copyright © United Nations, 2019
All rights reserved worldwide
United Nations publication issued by the Economic Commission for Europe

Preface

Migration shapes societies. Its economic, social and demographic impacts are large and increasing. Policymakers, researchers and other stakeholders need data on migrants – how many there are, their rates of entry and exit, their characteristics, and their integration into societies. These data need be comprehensive, accurate and frequently updated. There is no single source that can provide such data, but by combining several sources together it might be possible to produce the information that users need. Some countries have developed methods for combining administrative, statistical and other data sources for the production of migration statistics.

This publication, developed by a task force of experts from national statistical offices and endorsed by the Conference of European Statisticians in June 2018, provides an overview of the ways in which data integration is used to produce migration statistics, based on a survey of migration data providers in over 50 countries. Thirteen case studies provide more detail on data integration in various national contexts. The publication proposes principles of best practice for integrating data to measure migration.

The publication is designed to guide national statistical offices and other producers of migration statistics, while also offering users an insight into the production of the migration data they use.

Acknowledgements

The present guidance was prepared by the UNECE Task Force on Data Integration for Measuring Migration, with contributions of the following members: Antonio Argüeso (Spain, Chair of the Task Force), Regina Fuchs (Austria), Enrico Tucci (Italy), Monica Pérez (Italy), Kirsten Nissen (New Zealand), Marcel Heiniger (Switzerland), Jay Lindop (United Kingdom), Paul Vickers (United Kingdom), Jason Schachter (United States of America), Stephanie Ewert (United States of America), Anthony Knapp (United States of America), Giampaolo Lanzieri (Eurostat), Nathan Menton (UNECE), Paolo Valente (UNECE). The following national experts contributed the case studies for their respective countries: Mélanie Meunier (Canada), Adam Dickmann (Hungary), Ahmad Hleihel (Israel), Jeroen Ooijevaar (Netherlands), Baiba Zukula (Latvia). The contribution of all these experts is acknowledged with much appreciation.

Table of Contents

Preface	iii
1. Introduction	1
2. Definition of “data integration”	3
2.1 General features	3
2.2 Working definition	4
3. Survey on data integration practices	6
3.1 Overview of the main results of the survey	6
3.2 Recommendations	7
3.3 Summary of results	7
4. Case studies	8
4.1 Case studies of countries with population registers	8
4.1.1 Austria	8
4.1.2 Hungary	10
4.1.3 Israel	12
4.1.4 Italy	14
4.1.5 Latvia	17
4.1.6 Netherlands	19
4.1.7 Spain	21
4.1.8 Switzerland	25
4.2 Case studies of countries without population registers	28
4.2.1 Australia	28
4.2.2 Canada	29
4.2.3 New Zealand	31
4.2.4 United Kindgom	36
4.2.5 United States of America	39
4.3 Summary	40
5. Metadata	43
6. Conclusions	45

7.	Recommendations	46
7.1	Improve access to administrative data for national statistical offices	46
7.2	Use administrative data for migration statistics.....	46
7.3	Combine data from different sources using ‘presence signals’	46
7.4	Pay attention to the quality of integrated data.....	46
7.5	Be transparent about data integration methods used and develop standards	47
7.6	Promote international comparison and exchange of migration data.....	47
8.	Further work.....	48
8.1	The potential of big data.....	48
8.2	Longitudinal data	48
9.	References.....	49

List of Figures

Figure 4.1	Data sources for annual Hungarian population calculation	11
Figure 4.2	Monthly presence scheme of continuity patterns in job and study activities	16
Figure 4.3	Foreign population in Spain according to the population register 1998-2010 (millions)	22
Figure 4.4	New Zealand usual resident count of overseas born population	34
Figure 4.5	Percentage overseas born of New Zealand population	34
Figure 4.6	Calculation of Long-Term International Migration	38

List of Tables

Table 4.1	Process scheme and counts of population groups (in thousands) according to their likelihood of being usually resident in Italy	16
Table 4.2	Doubtful records in the 2011 pre-census file. Selected nationalities	24
Table 4.3	Doubtful population in the pre-census file group by nationality-province-age	25
Table 4.4	International migration data collections and processing	33

1. Introduction

1. Data integration has been a topic of great interest in recent years, especially as data needs have grown under constrained fiscal climates, combined with increased concerns about respondent burden and privacy. The UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS) undertook a project on data integration in 2016-2017. The project included experiments with data integration in various areas and led to the development of a practical online guide to data integration for official statistics¹. Eurostat has supported research on methods to improve data integration, such as the ESSnet project in the area of Integration of Survey and Administrative Data². This project was an initial attempt to create a common methodological basis for integrating different data sources.
2. Many countries have an interest in data integration specifically for the purposes of measuring international migration. Given the challenges in collecting migration statistics, it is often useful to collect data from several different sources. Data integration can reduce coverage or accuracy problems that may concern overall migration stock and flow data, and can enhance the richness of migration data by adding socio-demographic or economic dimensions to existing data.
3. At the European level, the importance of data integration to improve migration statistics was reaffirmed at the 2017 annual Conference of the Directors-General of National Statistical Institutes, which included a statistical session on 'Population Movements and Integration Issues - Migration Statistics'³. The 'Budapest Memorandum' discussed at the conference and approved by the European Statistical System Committee includes among its action points: *"To support the identification, assessment and adoption of new methods and data sources, particularly the increased use for statistical purposes of administrative data sources of appropriate quality ensured through ongoing quality assessment – either single registers, linked data from several administrative sources or combined with survey sources, and the opportunities offered by new data sources (e.g. Big Data)"*.
4. Methods to integrate different administrative data sources within a country (e.g. population registers, health, work, taxation, or education records) to supplement information missing from various sources (e.g. to measure outmigration for those who fail to deregister) have proved successful in some countries. In other cases, non-administrative data sources can provide information that is missing from administrative sources (e.g. use of household surveys to collect information not included in registers). 'Big data' from utility and telephone companies are also being considered in some countries as supplementary information. Other uses of data integration could be in reconciling different migration figures as derived from different sources, particularly in estimates for 'hard-to-count' migration populations, such as irregular migrants or emigrants.
5. Noting the importance and timeliness of this topic, in 2014 the UNECE-Eurostat Work Session on Migration Statistics discussed the utilization of different administrative data sources to measure international migration. The discussion focused on the opportunities and challenges faced when using administrative data, particularly given the differing levels of development of administrative data systems across the region. Ways to improve cooperation between national migration services, statistical agencies, agencies in charge of registers, and other producers of administrative data were considered, as were methods of integrating various administrative data sources to improve the measurement of migration. Participants also emphasized the need for practical advice and guidance on the production of metadata to facilitate comparisons between migration estimates produced by different countries.
6. The Work Session concluded by recommending further methodological work on the topic of integration of multiple data sources for measuring international migration, including data sources within a given country and between different countries. The work should include identification of good practices in communication between national statistical offices and producers of administrative data.

¹ <https://statswiki.unece.org/display/DI/Data+Integration+Home>

² https://ec.europa.eu/eurostat/cros/content/data-integration_en

³ 103rd DGINS Conference (Budapest, 20-21 September 2017). Papers and presentations are available at: <http://www.ksh.hu/dgins2017/presentations.html>

7. Based on these discussions, the Bureau of the Conference of European Statisticians (CES)⁴ established the Task Force on Data Integration for Measuring Migration in October 2015. The Task Force was mandated to prepare a set of guidelines and a description of good practices for integrating different data sources to improve the measurement of immigration, emigration and net migration. This publication presents the results of the work of the Task Force.
8. The publication provides a general overview of this subject as well as an overview of the types of data integration that are already in use in various countries – whether disseminated through regularly-published data or as part of pilot studies. Principles of best practices based on this overview are also provided to serve as guidelines for improving data integration for measuring migration in different countries. Country experiences are documented in this publication on the basis of a survey of migration data providers in over 50 countries, as well as on more detailed case studies from several countries.
9. Data integration is addressed in this publication by examining both **macro-data integration** – comparison/statistical modelling based on data which are aggregates (statistics) of individual-level records – as well as **micro-data integration** – the integration of data based on linkage/matching of individual-level records. Differing levels of overlaps of variables and/or individuals between different sources are also analyzed.
10. Further details on the above concepts are provided as part of the conceptual background and definition of data integration given in Chapter 2. Chapter 3 provides an analysis of the results of the survey of migration data providers. Practical examples of data integration in migration are shown in Chapter 4 through in-depth case studies from a number of countries. Chapter 5 presents a proposal for metadata to be considered for international migration statistics. Finally, the conclusions and general recommendations based on the survey and case studies are provided in Chapter 6.

⁴ The Conference of European Statisticians is composed of national statistical organizations in the UNECE region (for UNECE member countries, see www.unece.org/oes/nutshell/member_states_representatives.html) and includes in addition Australia, Brazil, Chile, China, Colombia, Japan, Mexico, Mongolia, New Zealand and Republic of Korea. The major international organizations active in statistics in the UNECE region also participate in the work, such as the statistical office of the European Commission (Eurostat), the Organization for Economic Cooperation and Development (OECD), the Interstate Statistical Committee of the Commonwealth of the Independent States (CIS-STAT), the International Labour Organization (ILO), the International Monetary Fund (IMF), the World Trade Organization (WTO), and the World Bank.

2. Definition of “data integration”

2.1 General features

11. The measurement of migration has long proved to be a very challenging task. Although considerable efforts have been made recently in both the academic and official statistics domains to improve the quality of migration statistics, there are still important margins of uncertainty about their accuracy. More recently, in a wider statistical context, attention has been given to the opportunities offered by ‘data integration’ as a potential additional action for the improvement of statistics (e.g., Eurostat, 2009, 2013; ESSnet, 2008, 2011, 2014; UNECE-HLG MOS 2017).

12. ‘Data integration’ is often mentioned as one of most effective approaches for enriching and improving statistical information. Various activities and projects over recent years have dealt with data integration, and these have adopted several different definitions. For instance, according to the Generic Statistical Business Process Model (GSBPM), data integration is an activity in the statistical business process in which data from one or more sources are integrated. In the HLG-MOS project mentioned above, data integration is defined as “the activity when at least two different sources of data are combined into a dataset”⁵. Yet, to the best of our knowledge, there is no internationally-agreed definition of data integration in the statistical community. There are various features which may help to characterize data integration, such as number and typology of data sources, methodology, timeliness, response burden, and not least its purpose.

13. Intuitively, integrating requires the availability of at least two inputs. These inputs are the data sets, i.e. organized information derived from selected data sources by means of a statistical operation. In the case of migration, data sets can be classified into three categories: (data sets derived from) statistical surveys (including exhaustive surveys, i.e. censuses), administrative registers, and big data. These data sources can be integrated in all possible combinations and the process can be repeated, generating data sets of mixed nature which can in turn be integrated with other data sets.

14. The way data sets are integrated depends mainly upon whether the information they contain is on the micro or macro level and, for individual records, upon the availability of a common identifier. For integration at the micro level, this latter feature determines the choice between statistical matching and record linking, the latter usually being applied when an identifier is available in all data sets being integrated. For integration at the macro level, i.e. between data aggregated from single records, the process should take place in two steps: first, cleaning necessitated by differences in concepts and operational definitions; second, reconciling (‘balancing’) the data, using various statistical techniques or simply calling on expert opinion.

15. In the case of migration data, the most common individual identifier is the Personal Identification Number (PIN). In fact, micro-level integration is preferred for the setup of population registers, of which migration statistics are often a by-product. It should be noted that a PIN is not the only possible common identifier, as different identification coding can be used to link data related to sub-national aggregations, administrative entities, file records, and so on.

16. Many data sources can be used as inputs to a population register, and in some countries additional data sources have been integrated over time, enriching the statistical information on the population and, indirectly, on migration. This can also be the case for registers of the foreign population or alien registers, although usually to a lesser extent. Data integration can thus be seen as an expanding process, where other integrated data sets are added to the first following lessons learned, refinement of integration methodologies, and access to new data sources.

17. Another distinctive feature of the measurement of migration is the fact that data sources other than national ones can be used, since migration is a phenomenon experienced by two countries at a time (receiving and sending countries,

⁵ <https://statswiki.unece.org/pages/viewpage.action?pageld=169018059>

or hosting country and country of birth, or whatever variable is used for identifying migrants). Exploitation of this intrinsic international feature is limited mostly to the exchange of aggregated data, since concerns about privacy and national security can lead countries to oppose requests for exchange of international migration microdata. Examples of such exchange do, however, exist in the Nordic countries, where international microdata exchange is used to improve the quality of the population registers.

18. Adding new data sources may lead to changes in the timeliness of the statistical output, i.e. changes in the delay between the reference period and the availability of results. While the general rule is that the timeliness of a specific output from integrated data is given by the data source with the larger delay, statistical modelling may actually reverse this situation by giving the opportunity to release more timely estimates. It is likely that a worsening of the timeliness would only be accepted when the data integration brings a relevant improvement of the coverage or accuracy of the statistical measurement.

19. The reduction of response burden is sometimes mentioned as one of the opportunities offered by data integration. To date, most of the uses of alternative data sources for migration statistics have involved replacing one data source with another that has a lower burden on the respondents or data providers, or improving the statistical operation applied to the data source. From a conceptual perspective, this may be seen as an enhancement of efficacy rather than an enrichment of data availability, and therefore not exactly unique to data integration.

20. Inputs from alternative data sources may also be used for the sake of validating the statistical output of a single official data source. This can be seen as **comparison** rather than **integration**. This may also be the first step in a process of progressive integration of these data sources, especially when the additional data sources do not comply with the requirements of official statistics.

21. Another case is when each data source covers a specific population (migrant) sub-group. The pooling of these data to produce an overall statistical output could be labelled as 'compilation' rather than true 'integration', since there is no actual merging at the level of the single record (either individuals or aggregated entities).

22. From the considerations above, it is clear that a defining feature of data integration is certainly the use of multiple data sets, but further to this, whether or not an activity qualifies as data integration is conditional upon the way that the data sets are used together. Some statistical operations border on data integration, such as data compilation and data comparison. As for the other features of data integration, the level of detail of the data matters for the methodology to be applied, rather than for the identification of a data integration activity. Timeliness cannot be used as a criterion, since data integration may in principle lead either to an improvement or a worsening of the delay of the results, and timeliness improvements can also be achieved by applying statistical operations other than integration. Likewise, any generic reference to quality improvement of the statistical output (including reduction of response burden) would not help to provide a definition, given that any statistical operation should in principle aim to have such an effect. Obviously, this does not imply that data integration does not have such positive effects, but only that these are not useful criteria to identify distinctive features of data integration.

23. The outcome of a data integration activity should be an 'enriched' or 'higher quality' data set. Thinking in terms of a generic data matrix $n \times p$, with n records and p variables, such enrichment could take place in both directions, improving the coverage (i.e., increasing n) and/or the information on the same records (i.e., adding new variables to the p set). As for the higher quality, this would not be reflected in changes of the size of the resulting data set, but rather in its content (e.g., in the weights of a sample survey). On the other hand, data integration could potentially decrease coverage, if some type of migrants are included on one dataset, but not on another. For example, irregular migrants who might appear on a survey, but not in an administrative database. This is especially true if "unmatched" cases are dropped.

2.2 Working definition

24. One available definition of data integration comes from the information technology (IT) environment (SDMX, 2009): *"The process of combining data from two or more sources to produce statistical outputs."* In the same reference, it is also clarified that *"Data integration can be at the micro-level, where it is often referred to as matching, or at the macro-level."*

25. The definition above is based on inputs (“*two or more sources*”) and purpose (“*to produce statistical outputs*”). The latter is rather generic, because a statistical output can be seen as the end of any statistical process and it does not necessarily require data integration.

26. For the purposes of this document, the following working definition is proposed: “*Data integration is a statistical activity on two or more data sets resulting in a single enlarged and/or higher quality data set.*”

27. In the course of integration, data can be processed at two main levels: micro- and macro-data integration⁶, defined as follows:

- ‘Micro-data integration’: the integration of data based on record linkage or statistical matching of individual-level records using key identifying variables;
- ‘Macro-data integration’: the combination of data based on aggregates (statistics) of individual-level records.

28. The reference to a generic ‘statistical activity’ leaves the definition open to include the use of any methodology, either on the macro or the micro level, including the use of expert opinion. It also encompasses the work on conceptual differences, a fundamental step in any statistical process. Limiting the definition only to those processes involving statistical matching and/or record linking would have excluded the macro-level integration.

29. “*Two or more data sets*” highlights the multidimensional nature of data integration. It also indicates that, while integration ends in a single outcome, it is an activity that needs to be repeated each time such outcome is needed. In other words, for the regular production of statistics from more than one data set, integration is not an occasional, once-and-for-all, statistical activity. The ‘integrated’ result must be generated anew each time a fresher data input becomes available⁷.

30. The use of the term ‘data sets’ instead of ‘data sources’ points to the fact that the subjects of integration are actual data, and not the data sources from which they are derived. For instance, merging two existing sample surveys into a single encompassing survey by modifying the questionnaires, the sampling design, etc., is not considered **data** integration. It also supports the point made above that the integration does not transform the input data sources into new, mixed data sources: it is the **outputs** of those data sources that are the core of the activity. The integration operation is thus likely to be as systematic as the production of statistics from the data on which it is applied.

31. The outcome of data integration is first and foremost a ‘single’ data set: a new set of data where all the information from the input data sets is reorganized in a harmonized fashion. This is not simply the union of multiple data sets, since redundancies, conceptual differences, and any other source of bias should be properly treated. This is an ‘enlarged’ data set, meaning a structured new set of data whose dimensions cannot be smaller than the largest corresponding dimensions of the individual input data sets, and/or a ‘higher quality’ data set, whose elements have been changed due to data integration with a (possibly measurable) increase in the statistical quality.

⁶ In some contexts the term ‘micro-macro’ is used when micro- and macro-level data are combined. This approach can be considered as a case of macro-data integration

⁷ Identifying conceptual differences in a context of data integration is possibly an exception to the ‘repetition rule’, because such activity would need to be carried out only once, unless the conceptual framework of the input data changes over time.

3. Survey on data integration practices

32. In order to collect the information needed to carry out its work, the Task Force carried out a survey on data integration practices. A questionnaire was designed to collect information on current practices in the use of combined or integrated data sources for the measurement of immigration and emigration flows, and for outputting statistics of the migrant population and the foreign-born population.

33. The questionnaire was sent to National Statistical Offices (NSOs) of UNECE member countries⁸ in September 2016. Fifty-six countries provided responses to parts of the survey relevant to their data collection practices.

34. The results of the survey were analyzed between the end of 2016 and the first quarter of 2017. The main results are summarized below.

3.1 Overview of the main results of the survey

35. Some form of data integration is essential for enabling the production of statistics on international migration flows. Ideally, data integration is done at the individual-record level, but in practice this is achievable mainly for countries with complete population registers that account for exits of nationals and entries of foreign nationals.

36. For many countries the actual migration event is not registered by one single source. Statistics on migration flows can only be derived by merging contributing components that account for changes in the country's resident population due to migration.

37. Statistics on international migration flows can be produced using a single source in cases where it is possible to maintain an administrative collection of the migration event (actual border crossings) or where a total population register-based system exists. However, quality assurance and further disaggregation of these statistics may require integration with supporting data sources. This may simply be in the form of an adjustment of migration flows following comparison with available statistics from other countries (mirror statistics). Grouping of individuals in a single source may be required for longitudinal observations of border movements as a way of classifying migrant status.

38. The survey results highlighted an extensive variety of uses of data integration. In some countries population registers of nationals may not be linkable at the unit record level to a separate administrative collection on foreign nationals. In such cases the population register may serve as a source for cross-checking and imputation of missing data.

39. A population census or a population register are the most common sources for producing statistics on the foreign-born population. The population census may serve as a base population, and subsequent integration with an administrative collection enables the updating of statistics on the foreign-born population.

40. When multiple sources are used to inform the migration statistics, some countries reported a need for assessment of the sources in order to prioritize in terms of timeliness, completeness, reliability, coherence and other quality dimensions.

⁸ In addition to the 56 countries that are members of UNECE, the questionnaire was also addressed to other countries that participate regularly in the activities of the Conference of European Statisticians, including Australia, Chile, Colombia, Japan, Mexico and New Zealand.

3.2 Recommendations

41. As countries advance and standardize registers and social and population statistical data sets, data integration is more likely to take place at the unit record level across multiple sources. Greater attention to quality assessment of the integrated data will necessarily become an integral part of the production of international migration statistics.

42. It is increasingly important for statistical offices to provide a transparent presentation of the rules and methods used for the data integration processes that are used to produce statistics on international migration flows and other types of international migration statistics. This creates opportunities to develop methodological standards for data integration practices across national statistical offices in the future. Individual countries will always have their own collection and data integration practices, but there are likely to be more technical forums opening up for discussion of the rules and methods used in the combining of sources for the production of migration statistics.

3.3 Summary of results

43. Nearly all respondent countries produce statistics on international migration flows, and more than half of these countries would use a combination of data sources. Some countries would use a central register of all registrations and de-registrations of residents, and integration of the source at the individual record level with administrative collections of emigration and immigration flows, and deaths, supports a refined production of migration flows. Other countries integrate population registers with survey information or other administrative collection of non-nationals at the macro level.

44. Overall, about one third of respondent countries (35 responses) integrate data sources at the macro level using statistical modelling or other methods to combine aggregated data, and a slightly smaller proportion of countries apply integration methods at the unit record level. However, it should be noted that a few countries used other sources for cross-checking and complimentary information only. About half of respondent countries noted observation units from the integrated data sources were partially overlapping; a smaller proportion had identical observation units when combining sources; and a few reported the use of mutually exclusive observation units.

45. Less than half of countries prepare statistics on international migration flows from a single source. Statistics on migration flows from a single source can frequently be produced when it is possible to maintain an administrative collection of all border crossings or when migration flows can be derived from a population register. Another less frequent option is the use of the population census to estimate migration flows indirectly from population stock measures.

46. Around 90 per cent of respondent countries produced statistics or collected data on the foreign-born population. Just over half of these countries use more than one data source and usually by data integration at the unit record or macro level. The main types and features of the sources used, either single or integrated, were a five-yearly population census, population register, survey, or administrative collection. The main purpose of integration was reported to be for production of statistics on the foreign-born population.

47. Around 60 per cent of respondent countries reported that they did not produce other statistics on migrants or the foreign-born population. A few countries reported that integration of data sources allowed the preparation of statistics on topics such as asylum seekers and grants, residence permits, work permit holders, irregular foreign workers, socio-economic characteristics of migrants, and reason for settlement.

4. Case studies

48. Taking into account the information collected in the survey on data integration practices, the Task Force identified a number of countries with significant experience in data integration. These countries were invited to provide a short text describing the primary migration data sources, the reasons why data integration is considered necessary, the methods used and the benefits of integration.

49. The texts provided by countries are presented in this chapter as a collection of significant practical examples of data integration for migration statistics. The case studies presented for the various countries differ significantly because the national contexts vary with regard to a number of significant aspects, such as: availability of population or other registers; existence of a personal identification number; the possibility of linking data across registers; legal frameworks; and public sensitivity concerning the use of register data. Therefore, it is very difficult to classify the case studies or group them into categories. However, a very significant aspect with regard to data integration – which influences heavily the methods adopted – is whether or not a population register is used in the country. Therefore, the case studies of the countries with systems that are mainly based on a population register or other administrative registers are presented first, followed by the countries with systems without population registers, based mainly on passenger cards or administrative data other than registers.

50. The majority of the case studies presented follow a structure suggested by the Task Force, including the following sections:

- a) Original/primary source of data for migration statistics;
- b) Limitations of the original or main source. Why is integration necessary?;
- c) Methods used for data integration;
- d) Benefit of integration with statistics from other sources. Selected case studies follow a different structure depending on the information available or the specific context of the country.

4.1 Case studies of countries with population registers

4.1.1 Austria

51. Since 2002, Austrian migration statistics have been based upon data from the Central Register of Residence ('Zentrales Melderegister' (ZMR)). Residents who establish their home in a private or institutional household have a legal obligation to register and de-register their residence within three days of moving. All residents, independent of their nationality⁹ or length of stay must register if their stay exceeds three days¹⁰. Statistics Austria receives and processes all residence registrations and de-registrations on a quarterly basis.

52. Data from the Central Register of Residence are enhanced by integrating data in two separate processes. First, information on deaths from the Organization of Austrian Social Security ('Hauptverband der Sozialversicherungsträger' (HV)) is linked to the residence register at the unit record level, adding information on deceased persons and their missing de-registrations. Second, in order to adjust for missing de-registrations of emigrants, yearly estimations identify potential nominal members employing information from several administrative data sources. While both steps link data on the micro level in the final stage, the underlying processes differ substantially.

⁹ Including asylum seekers and refugees

¹⁰ Foreign diplomatic personnel are exempt from registration

4.1.1.1 Adjusting for uncounted deaths in the population register

53. Data from the Organization of Austrian Social Security serve as a supplementary data source for improving the quality of Austrian population and migration statistics. The Organization of Austrian Social Security is the umbrella organization of all social security funds in Austria. It collects information on all persons insured in Austria and their dependents. Their database coverage is determined by insurance status, rather than residence registration status¹¹. Therefore, the residence register and the social security data base have partly overlapping populations, but also cover categories of persons that are represented only in one or the other database.

54. Registration of deaths in the Central Register of Civil Status (ZPR) is compulsory for all deaths occurring on national territory, as well as deaths of Austrian nationals occurring abroad. Following the death registration, the deceased is de-registered from the residence register. However, deaths of Austrian residents of foreign nationality, occurring on foreign territory, are not subject to registration. The notice of death may reach the Austrian authorities with a delay.

55. The social security database thus supplies complementary information on deaths, and is employed to refine the total population and migration count. These data are matched with the migration flows from the residence register via bPK, an anonymized 27-digit key that allows the connection of information from different administrative data sources.

4.1.1.2 Estimation of potential nominal members in the population register

56. At the reference date of register-based censuses (the most recent being 31 October 2011) people appearing only in the Central Register of Residence, but in no other administrative register, are identified as suspicious cases necessitating further investigation. They are contacted to confirm their presence in Austria. Those not having responded are seen as nominal members and therefore excluded statistically from the population. Municipalities are also informed of the results and are asked to clean up their administrative records.

57. For the inter-censal period a similar exercise is undertaken annually. In this case, people are not contacted directly to confirm their presence, but rather a share of those identified as unique cases in the Central Register of Residence are statistically excluded from the total population. The likelihood of being a nominal member is determined by applying a logistic regression model¹².

58. The results of this exercise are integrated into quarterly population statistics once a year (upon publication of the final results). Deviations between quarterly population statistics and those identified as nominal members by the so-called mini register-based census are integrated in two ways:

- a) By relocation abroad of non-recognized registered residents; and
- b) By relocation from abroad of previously non-recognized residents.

59. The population adjustment for nominal members thus has an impact on the migration flows in a direct way. Potential nominal members not recognized in population stocks are assumed to have relocated abroad (and thus statistically counted as emigration). In turn, previously non-recognized residents can re-join the population stock if they show presence signals in other administrative records than the residence register in the subsequent year. In this case, a statistical record for re-entry to Austria (immigration) is created.

60. In contrast to directly linking information from a single other data source at the micro level (as in step “(a) Adjusting for uncounted deaths in the population register” above), a different methodological approach is used: first, several data sources are consulted to identify registered residents who are potential nominal members, and second, statistical estimation techniques finally select the individuals who will no longer be recognized residents. Here, data linkage on the unit record level creating statistically-produced emigrations and immigrations is only a downstream process in the estimation of nominal members from various administrative records.

¹¹ The social security database covers, for example, commuters from abroad without residence in Austria, but does not supply any information on persons who have never been insured, received pension payments, child support or any other social service.

¹² Detailed documentation in German: http://www.statistik.at/wcm/idc/idcplg?IdcService=GET_PDF_FILE&RevisionSelectionMethod=LatestReleased&dDocName=073537

4.1.2 Hungary

4.1.2.1 Original/primary source of data for migration statistics

61. At the Hungarian Central Statistical Office (HCSO) the Population Census and Demographic Statistics Department produces the annual population figures, which are based both on registers of vital statistics and on migration data.

62. The number of foreigners with residence or a settlement document as well as those with refugee status and persons under subsidiary protection, who have a registered address in Hungary, are calculated from the administrative registrations.

63. 4.1.2.Migration data are produced using various types of registers. These data sources are under the scope of various authorities, as presented below.

4.1.2.1.1 Immigration and Asylum Office (IAO)

64. A memorandum of cooperation has existed between the HCSO and IAO since 2009. The Office of Immigration and Asylum is responsible for two different types of registers:

- a) EEA-register: the register contains data on EU and EEA citizens and their accompanying family members;
- b) IDTV-register: this register contains data on third-country nationals.

65. Both registers are established for administrative purposes. The statistical data collection is a 'by-product' of the registers.

66. Using the registers of IAO, two types of foreign migrants can be distinguished:

- Non-national immigrants are foreign citizens who applied for and were granted a residence permit by the Immigration and Asylum Office, and who hold a registration certificate (residence card in the case of EEA citizens, or residence permit in the case of third-country nationals);
- Non-national emigrants are persons who used to own a document which entitled them for a legal stay in the territory of Hungary, but this document expired or their residence permit was invalidated. From 2012, their number includes estimations based on the Census of 2011.

4.1.2.1.2 Ministry of Interior

67. The Ministry of Interior runs the central population register (CPR). The CPR contains data on persons immigrating to Hungary and data on persons emigrating from Hungary who are obliged to register an address in the CPR (both Hungarians and foreigners; according to the current Hungarian law about two-thirds of foreign citizens in Hungary are present legally).

68. In the CPR data on persons for whom international protection is granted and data on persons naturalized in Hungary are stored. A person naturalized in Hungary is someone who became a Hungarian citizen by naturalization (he/she was born as a foreign citizen) or by re-naturalization (his/her former citizenship was abolished). The rules of naturalization in Hungary were modified by the Act XLIV of 2010. The act introduced a simplified naturalization procedure from 1 January 2011 onwards, and made it possible to obtain Hungarian citizenship without residence in Hungary (for persons who can prove that they have Hungarian ancestors).

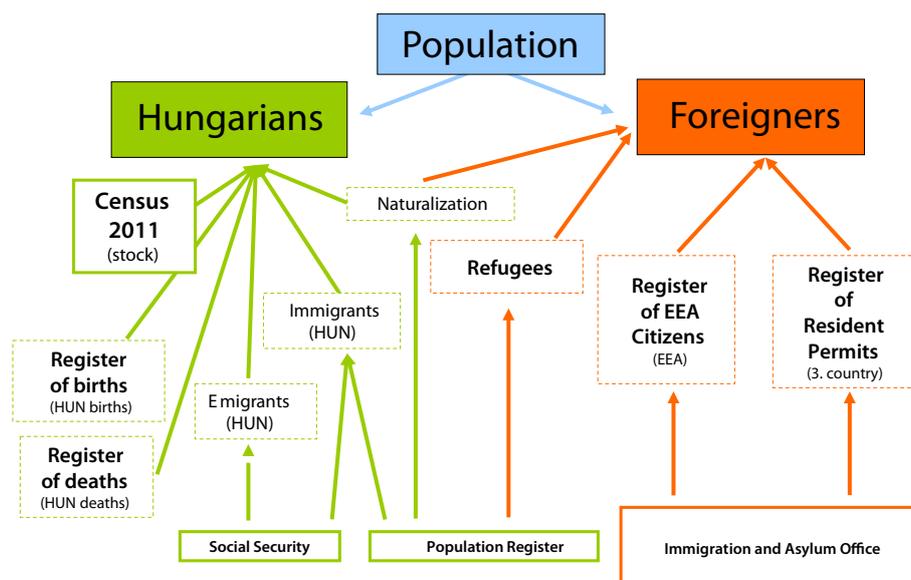
69. The data published by the HCSO refer only to those new Hungarian citizens who have an address in Hungary.

4.1.2.1.3 Ministry of Human Capacities: National Health Insurance Fund of Hungary (NHIF)

70. The NHIF is also one of the main data source on migrants who are insured in Hungary, which is the vast majority of the migrant population. For calculating the migrant population, the data of NHIF are used to establish the number of immigrations and emigrations of Hungarian citizens (the law states that Hungarian citizens must deregister from the NHIF if they leave the country for another EEA member state, and that they must re-register if they return to Hungary from a stay in an EEA country). Migrants with Hungarian nationality can thus be divided into the following categories:

- National emigrants: persons who left Hungary and reported it to the National Health Insurance Fund;
- National immigrants belonging to one of two subgroups:
 - Hungarian citizens who returned from a period abroad and reported it to the National Health Insurance Fund;
 - Hungarian citizens who established a Hungarian address (and registered in the Population Register) after having been granted Hungarian citizenship while living abroad (without previously having Hungarian residence).

Figure 4.1 Data sources for annual Hungarian population calculation



71. The Census 2011 is highlighted in the figure because the administrative data of 31 December 2011 was harmonized with the Census 2011 results, and the data from subsequent years are based on this harmonization.

4.1.2.2 Limitations of the original main source. Why is integration necessary?

72. The Hungarian migration statistics system described above is a very complex system because the relevant data on migration are stored in various registers and in some cases even if the registers contain migration-related data they are stored in various sub-registers. This is the case, for example, for data on foreigners.

73. In order to produce the numbers of immigrants and emigrants (both of Hungarian citizens and of foreigners) it is necessary to integrate the data on migration from the different data sources.

74. It must be emphasized that in the Hungarian case there is no source that can be identified as the 'main' data source. The data sources are almost equally relevant for the production of migration-related statistics: that is, the data sources are complementary to each other.

4.1.2.2.1 Production of migration data (Hungarian citizens and non-nationals)

75. For the production of flow and stock data on migrants in Hungary, the HCSO uses the data from the registers of the Immigration and Asylum Office (EEA-register and IDTV-register). During the procedure these two databases, which are provided by the Immigration and Asylum Office, are merged first, in order to clean the data and remove double registered records. The data are linked to the CPR as well – it is essential to note that the CPR contains data on foreigners who have address registration [third-country nationals with residence permits (short-term stay) are not obliged to register in the CPR]. Second, the registers of OIA are merged with the CPR because the data of persons taken under international protection are registered in the CPR.

76. The cleaning of the primary data sources is an important step in the procedure because the primary data source contains numerous double-registered data.

77. In order to get the migration data of Hungarian citizens HCSO currently does not use any data integration procedure. The data source for flows of Hungarian citizens is only the NHIF.

4.1.2.3 Methods used for data integration

78. The IAO register data and the data of the CPR are provided twice a year to HCSO (at the beginning of the year and at the end of the year). The data is provided at the record level and the records are linked together by a key-variable developed using SAS. The linking method is a deterministic linking via unique identifiers, which enables high quality integrated data sets. Of course, this integration method imposes high demands on the administrative data infrastructure. Mismatches occur from misspelling and name changes.

79. First the data of IAO are cleaned to remove double-registered data, and afterwards the cleaned data set is interlinked with CPR data.

4.1.2.4 Benefit of integration with statistics from other sources

80. Integration of migration data is a necessity for HCSO because even the relevant data maintained by a single authority are stored in different registers in Hungary. The motivation for data integration arises from a particular shortcoming of administrative data, but the resulting integration creates additional information.

81. On the other hand, this data integration enables the production of high quality data. Data integration is necessary for the production of migration-related data and it is also important for the annual population calculation.

4.1.3 Israel

4.1.3.1 Original/ primary source of data for migration statistics

4.1.3.1.1 For population estimation purposes

82. The Israeli Population and Immigration Authority (PIA) is responsible for border control registration (BCR) and for registration at the population registry (PR). The PIA provides the Central Bureau of Statistics (CBS) with separate files for Israelis and for foreigners. These administrative data make it possible to distinguish between immigrants and non-permanent residents (NPR) and allow for measurement of the flows and the stock of emigrants and immigrants.

83. For NPRs, the files required to produce their estimates include their visa at the time of admission. Although the files include different types of visas, we only refer to visas for foreign workers. PIA provides the CBS with a special registration for NPRs who entered Israel illegally, not through an official border crossing.

84. For permanent residents, CBS uses the new registration at the PR, distinguishing between new citizens (mostly Jewish or of Jewish extraction who immigrate to Israel under the Law of Return) and non-citizens (mostly family reunification).

85. For emigration estimation the CBS uses the BCR to examine the length of stay abroad.

86. An 'out emigrant' is defined as someone who has gone abroad and stayed abroad for at least one year. A 'return emigrant' is defined as someone who has returned from a stay abroad of at least one year and who has resided in Israel for at least three consecutive months after returning from abroad.

4.1.3.1.2 For analytical characteristics

87. The permanent and non-permanent residents' data provided by PIA is the primary source of administrative data for analytical research and policy needs. This data source is a census of the immigrant population and, as such, offers the capacity to study small populations in detail. This data source on permanent residents contains socio-demographic information about immigrants who have arrived in Israel since 1970. Information about characteristics related to destination in Israel, country of origin, intended occupation or level of education is obtained at the time of admission. For additional administrative information we link PIA data with other administrative data (education, employment, income, etc.).

88. The Census of Population and all household surveys include a module of questions that allow us to know the year of immigration and the country of birth of permanent residents. This module enables the CBS to get information on all topics of the surveys.

4.1.3.2 Limitations of the original or main source. Why is integration necessary?

4.1.3.2.1 For population estimation purposes

89. The information available on non-permanently resident foreigners is the information received on the day of admission. This information does not include complete information on the place of residence. The information identifying the individual in these files does not allow the linkage of records with other sources (administrative or surveys). These limitations restrict the CBS in preparing the data beyond that which exists in the administrative files. There are also some cases where no linkage is possible between arrival and departure records. This may affect our stock estimation. This limitation exists especially for records of people with tourist visas, and limits us in the estimation of overstayers (unreported migration).

90. Many Israelis have citizenship of another country. In most cases, the BCR knows how to identify dual citizenship and registers their border crossings even if the person has crossed the border with a non-Israeli passport. In some cases, however, the BCR misses the information of dual citizenship. This limitation affects the estimation of emigration.

4.1.3.3 Methods used for data integration

4.1.3.3.1 For population estimation purposes

91. In the case of permanent resident immigration, information is integrated at the level of the individual record. All the files contain the same variables and are linked using a unique ID (personal identifier of PR). Measuring the number of immigrants entering Israel in a given time period is straightforward, and adjustments to the data are not required. Information is available for each person entering Israel with immigrant status from PIA administrative files. Every month, PIA prepares a data file containing the records of immigrants in the previous month and makes it available to the CBS, as well as any additions or updates to data already received.

92. For the NPR population, there are three major subgroups, which are administratively different: foreign workers with permits, irregular foreign workers and immigrants who cross the border illegally. Estimates of these three groups must be produced separately. In the case of foreign workers with permits and irregular foreigners we get a list from PIA and we use it as is. For more information about the methodology in the case of irregular workers, please refer to Sheps M. (2016).

93. As for emigration, please refer to Sheps M. and Hleilel A. (2006).

4.1.3.3.2 For analytical characteristics

94. To produce complete information or to carry out longitudinal studies, the CBS uses exact-matching record linkage techniques to combine the information on immigrants with other administrative files (such as tax files, education files, etc.), and with population censuses and surveys.

95. Matching is also used to add variables from data sources related to the day of admission.

4.1.3.4 Benefit of integration with statistics from other sources

4.1.3.4.1 For population estimation purposes

96. The use of a unique ID in all administrative files, and the collection of this ID in most Israeli surveys and in the population census, allow the CBS to combine all these sources together. The main benefit of this operation is to provide estimates that are of greater quality, which will be useful on the national level.

97. In addition, the integration of information enables the CBS to reduce the response burden and reduce costs of data collection.

4.1.3.4.2 For analytical characteristics

98. The CBS allows its employees and external users to conduct data processing and research using the combined data. Researchers receive access to data in research rooms located in the CBS buildings. There is no standard data file on immigrants that is accessible to researchers. Usually every researcher defines his or her data according to the needs of the research. The data allow users to research the absorption and integration of immigrants into Israeli society and in the economy, enabling the monitoring of their movements in the population. The family relationships recorded in the PR allow users to obtain information about the second generation as well.

99. In the case of emigration, studies have been performed on the characteristics of emigrants before departure and afterward. Linked data on emigrants from various sources permits the study of the 'brain drain' phenomenon, for example.

4.1.4 Italy¹³

4.1.4.1 Primary source of data for migration statistics: the population registers ('Anagrafi')

100. Since the early 1990s the Italian National Institute of Statistics (Istat) has been trying to make the best use of the available administrative sources to produce international migration statistics, especially the population registers ('Anagrafi'). This has resulted in the development of a large statistical documentation that captures the main socio-demographic features of migration.

101. Many changes have taken place in recent years, both administrative and in terms of statistical data management and collection procedures. The digitalization of population registers and changes in the law have improved the productivity of administrative systems and have reduced the administrative burden on citizens. Such changes have had large effects on statistical production and have permitted an improvement in data quality in terms of timeliness, coverage and consistency with other administrative sources.

102. In Italy, official statistics on international migration flows are based on local population registers under the responsibility of the Ministry of Interior, which is responsible for supervising local administrations and national security. Changes of residence data source provide the size of international migration flows, information on both the origin and destination of a single movement, as well as the main socio-demographic characteristics of migrants. Istat collects individual data from more than 8,000 municipalities every year. The data collected are not usable as they stand because data quality problems are present due to errors during data entry, missing information or other invalid data. The validated data are cross-checked against the data in aggregated form (demographic balance) and the inconsistencies are removed in order to produce unique, more accurate and consistent statistics as required by the Regulation of the European Parliament and of the Council on Community statistics on migration and international protection¹⁴.

4.1.4.2 Limitations of the original source. Why is integration necessary?

103. Although the quality has improved considerably over the last decade, the accuracy of this register is still a critical issue, due mainly to the difficulty of recording international emigration but also due to the discrepancies between the administrative concepts and the international definitions. These lead to three main limitations of the original source:

- Over-coverage: In Italy, nationals living abroad and foreigners who have left the country permanently or long-term should be removed from the population register. However, emigrants have little motivation to notify the authorities of their departure and may remain for years on the registers before being removed as 'deregistration ex-officio';
- Undercoverage: statistics refer to the 'de jure' legally resident population, and not to the 'de facto' one: the foreign population usually resident in Italy but without official residence is excluded from the statistics. Although coverage of the usually resident foreign population in Italy is monitored, there are foreign citizens who do not want to be recorded in the population register, especially nationals from the European Union

¹³ Prepared by the Directorate for Social Statistics and Population Census, Italian National Institute of Statistics (Istat).

¹⁴ 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. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32007R0862>

Member States. Moreover, those who no longer meet the requirements for staying in Italy and immigrants who entered without authorization cannot be recorded in the population register. The latter, who do not appear in any administrative sources, represent a hard-to-count sub-population and they require a suitable survey to be counted;

- Time criteria: the source does not include information on the intended length of stay, although it is likely that who registers (or deregisters) intends to change his or her residence for at least 12 months.

4.1.4.3 Methods used for data integration

104. A solution to the coverage challenges of population registers is the use of an integrated system of registers: official population registers could be linked together with other subject-specific administrative sources (related to labour, education, taxation, earnings, etc.). This system could be used to identify groups that correspond to the national or international definition of 'usually resident population'. Under-coverage in population registers could be detected by using individuals' signals of presence on Italian territory coming from other registers; whereas the absence of signals for people in the population register could be evidence of over-coverage.

105. Istat organized a Working Group with the aim of testing the use of administrative information for Census purposes and for producing population counts. The results of a trial carried out by Istat's researchers is summarized here.

106. Data integration is necessary to define a preliminary workflow for processing administrative data to derive usually resident population counts. Istat has built an integrated system of administrative sources to manage the increasing number of administrative datasets acquired for statistical uses and to maximize the benefits deriving from the huge amount of information available. To this end it was necessary to centralize some functions for acquisition, storage, integration and administrative data quality evaluation in the system called SIM (Integrated System of Microdata). This system manages social and economic administrative data on individuals and household characteristics (demographic aspects, employment status, level of education, place of residence, place of work and place of education). The integration step in the SIM is the process of linkage and physical integration of microdata coming from different sources: depending on the linking variables available, a suitable integration strategy and a set of algorithms are applied.

107. Integration means:

- Identifying each object (individual, economic unit) in the administrative data sources with an identification (ID) number that is unique and stable over time;
- Defining, for each object, the logical and physical relationships among sets of data coming from different sources.

108. SIM was created with the aim of supporting the statistical production process: the use of the SIM IDs enables the building of data structures that could constitute the starting point for statistical processes based on administrative data. To evaluate how administrative data could improve population counts, a thematic database of 'administrative signals' was created to support an initial experiment. Data used in the trial came from specific administrative sources already stored and integrated in the SIM system: population registers, permits to stay, employees and self-employed, compulsory education, university students, retired people, non-pension benefits, income and taxation.

109. By using a standardized time reference for the duration of an activity, it was possible to derive a monthly presence scheme. This feature consists of a sequence with a length of 24 digits, each of them representing whether or not a person was present for a specific activity and in which geographical area of the country. Clustering on these sequences, evident in the majority of the activity signals, showed the existence of patterns of continuity (Figure 4.2).

Figure 4.2 Monthly presence scheme of continuity patterns in job and study activities



110. Evaluation of consistency between the registers examined results in a classification of all the cases recorded in the experimental database, into groups that are useful when producing population counts. The process to derive these groups consisted of three steps (Table 4.1). The first step involved linkage between the experimental Population Register (PR), the activity signals from the Labour and Education registers (LE) and the Permits to Stay register (PS); in the second step, retired people or people with other benefit signals were identified among all the individuals without signals in the LE; finally, the Tax Returns Register (TR) provided indirect signals from people who could in some way prove their presence in Italy.

111. It is possible to organize these groups according to evaluation of usual residence. The individuals belonging to Group A, Group E and Group G could be considered ‘eligible’ as the usually resident population using the usual residence concept with 12 months’ time criteria. In contrast, Group C, Group D and Group H require a more thorough analysis (Table 4.1).

Table 4.1 Process scheme and counts of population groups (in thousands) according to their likelihood of being usually resident in Italy

Step I			Step II				Step III			
PR	LE	PS	Group	Count	RR-NPR	Group	Count	TR	Group	Count
61,068	37,704	3,378			20,764			26,649		
Signals					Signals			Signals		
Yes	Yes	-	A	36,618						
Yes	No	-	B	24,450	Yes	E	14,485	Yes	G	6,939
					No	F	9,965			
No	Yes	-	C	1,086						
No	No	Yes	D	351						

Table legend:

	Eligible as usually resident
	Uncertain residents

Administrative sources legend:

PR	Population Register
LE	Labour and Education Registers
PS	Permits to Stay Register
RR-NPR	Retired and Non-Pension Benefits Registers
TR	Tax Returns Register

112. At the end of 2012, about 3 million people were recorded in Group H which represents the sub-population recorded in the population register (PR) without signals from other sources. Almost 75 per cent of these people had Italian citizenship. Because of the absence of records in labor and education registers, Group H presents a high risk of unrecorded emigration that could bias population register data.

4.1.4.4 Benefit of integration with statistics from other sources

113. The integration of Italian administrative data sources proved to be very effective in exploiting the wealth of information already available. However, since signals can result from a temporary or occasional presence, it is necessary to carry out a characterization process by constructing derived variables which then make it possible to identify cases of permanent presence that correspond to the international definition of usual residence. In the trial, Istat defined a preliminary workflow to integrate the use of administrative sources and the official population registers, in order to calculate the usually resident population. Using this workflow, it is possible to identify a group of individuals eligible to be included in the usually resident population of Italy at a given reference date. This group of possible usual residents totalled 62.6 million individuals on 31 December 2012. From this group of individuals, it is possible to identify three main sub-groups:

- a) The subpopulation present in the population register, without signals from other sources (3.0 million individuals in 2012) that could have emigrated without giving notification of their departure;
- b) The subpopulation present in the population register that showed signals in other sources (58.1 million individuals in 2012);
- c) The subpopulation not present in the population register but with signals from other sources (1.5 million individuals in 2012) that could have immigrated without giving notification of their arrival.

114. The analysis of signal strength based on its continuity over time is only the starting point in the use of longitudinal data that can be processed with administrative sources. The main objective of the future trials should therefore be the study of longitudinal models, over several years, to produce subpopulation estimates which are more stable in relation to the fluctuations linked to the labour market, from which signals are derived.

4.1.4.5 References

Tucci E., Marsili M. and Terra Abrami V. "Improving quality of international migration outcomes by incorporating the micro-approach in managing current demographic accounting (MIDEA) and statistical population registers (ANVIS)", UNECE/Eurostat Work Session and Workshop on Migration Statistics, 10-12 September 2014, Chisinau, Republic of Moldova

Prati, Gallo, Chieppa, Tomeo et al. "Feasibility study of using usual residence concept with 12 month's criteria for all the breakdowns of population, births and deaths requested in the regulation (art. 8)", note prepared by the National Institute of Statistics of Italy for the REGULATION EU 1260/2013 on European Demographic Statistics

4.1.5 Latvia

4.1.5.1 Original/primary source of data for migration statistics

115. A population count is estimated on the basis of data files provided by Office of Citizenship and Migration Affairs (OCMA) from the Population Register sub-system of the Common Migration Information System. Individual data from these files include information on sex, birth date, country of, citizenship, legal marital status, code of administrative territory of declared place of residence according to the Classification of Territories and Territorial Units (CATTU), residence permit.

116. Since January 2013, the Central Statistical Bureau (CSB) has also been receiving information on registration of civil status (marriage, birth or death) from the Common System for Registration of Civil Status (CSRCS) sub-system of the OCMA Common Migration Information System. It receives this information as annual data files at the individual level (previously information was received from county civil registry offices in paper form). With the help of the above data, information is produced on the annual number of births and deaths.

117. Analysis of Population and Housing Census 2011 data allowed CSB to determine the population count and composition in the country. On 1 January 2011, Latvia's population amounted to 2,074.6 thousand people. As compared to the information published prior (according to the Register), the population count was 155 thousand or 7 per cent smaller. The data confirmed that part of country population does not fulfil requirements of the Population Register Law and information included in the register on population migration is incomplete.

4.1.5.2 Limitations of the original or main source. Why is integration necessary?

118. The following problems were identified within the population register:

- Non-deregistration or late deregistration (non-registered migration). Although according to the Population Register Law people are obliged inform the register about changes in their de jure marital status, children, usual place of residence etc., they do not always do it. According to the 2011 Census results, seven per cent of the population of Latvia had moved abroad, but had not changed usual place of residence in the register (according to the register, they resided in Latvia). Also, analyzing information from the population register a few years after the census, it is possible to see that there are quite a lot of people who correct the information in the register with a delay;
- Very old persons who are in the register but whose usual place of residence is not Latvia. In cases where such people have died abroad and the population register has not received this information, they remain in the register as alive. In the worst cases a person who has left the country without deregistration, and who therefore continues to have their place of usual residence recorded in the register as Latvia, may have died while abroad.

4.1.5.3 Methods used for data integration

119. In 2012 the CSB developed a new method for estimating population counts. The method is based on the population register and on individual data from other administrative registers (all data sources have ID codes which are used to merge data).

120. To evaluate the residence status of each individual, a logistic regression model was developed. With the help of administrative register data, more than 200 characteristic variables are developed for each person registered within the population register. The aim of the model is to predict the probability (number falling within the interval from zero to one) of being a resident for each individual, to estimate the status of the actual place of residence at the beginning of the year for every resident registered (at the individual level) in Latvia. The probability that is necessary to be included in the resident population differs depending on age and sex.

121. The model was developed using data from the 2011 Census on actual place of residence and data from administrative data sources in 2010, on 1 January 2011, or on 1 March 2011.

122. Every year a database on individual level is prepared with data referring to 1st January. The database permits the preparation of aggregate data tables for national and international data users.

123. Apart from OCMA data, population statistics are produced by using individual data from other administrative registers available to the CSB which correspond with the time period in question (starting from 2010). CSB has access to data from the following administrative registers:

- State Revenue Service (SRS);
- State Social Insurance Agency (SSIA);
- Ministry of Education and Science (MES);
- Agricultural Data Centre (ADC);
- Rural Support Service (LRSS);
- National Health Service (NHS);
- State Employment Agency (SEA);

- 32 out of 58 higher education institutions (89per cent of total number of students);
- Road Traffic Safety Directorate;
- Prison Administration.

124. CSB also has information from the Social Security Administration Information System (SSAIS) on persons who have received social benefits from local governments. However, this information is available only starting from 2012. These data are used to adjust the number of population in the 18-30 age group and to assess model quality.

125. Starting from 2016, the population count is specified by using SSIA data on persons who have been paid benefits or pensions and who are located in social care institutions. Additional analysis is also being carried out regarding the usual place of residence of the mother if the usual place of residence of a child (under 16 years of age) and father is Latvia, according to estimates.

126. International long-term emigration from Latvia to another country theoretically corresponds with international long-term immigration data from Latvia to the respective country. Such 'mirror statistics' are used to estimate international long-term emigration from Latvia.

127. When estimating international long-term emigration, information regarding immigration from Latvia is received from other countries, namely Denmark, Finland, Sweden, Norway, Spain, the Netherlands, Austria, Iceland and Germany. Not all countries prepare data on immigrants from Latvia, as Article 3 of the Regulation of the European Parliament and of the Council on Community statistics on migration and international protection stipulates that countries may provide information on immigrants with a grouped breakdown of previous place of usual residence. Hence the breakdown can be grouped into EU Member States; European Free Trade Association countries; candidate countries; other non-member countries.

128. To estimate emigration of the Latvian population to the United Kingdom and to Ireland, the following information is used: the number of UK National Insurance Numbers granted for the first time and the number of Ireland Personal Public Service Numbers granted for the first time. It should be noted that these data are used only to evaluate general trends, because these identification numbers are given to Latvian residents even if they stay in the UK or Ireland for less than one year.

129. Household survey data are used to evaluate the quality of estimation (individual-level data from the Labour Force Survey (LFS), the 'EU Statistics on Income and Living Conditions' survey (EU-SILC), the European Health and Social Integration Survey (EHSIS) and the European Health Interview Survey (EHIS). Evaluation of the results of this method was one of the tasks of the 2015 micro census.

130. Detailed information about the methodology is available on CSB's website (http://www.csb.gov.lv/sites/default/files/dati/demstat_metodologija_eng.pdf).

4.1.5.4 Benefit of integration with statistics from other sources

131. The main benefit of integrating different administrative data sources is to provide more precise estimates of population that will be useful not only for the statistical office, but also for national and international data users. The method developed will be used for the 2021 census which will be register-based.

132. The method has been presented and positive comments received from the Latvian Statistical Association, experts from the Central Bank of Latvia, and demographers. The main conclusion was that if the information on place of usual residence in the population register is of such low quality, the method developed is the best solution.

4.1.6 Netherlands

4.1.6.1 Primary sources for the measurement of migration statistics

133. Dutch demographic statistics are entirely based on Dutch population registers. As such, describing immigration statistics in the Netherlands basically boils down to describing the definitions and practices used in the population registers. However, some limitations are present. By integrating administrative data from many sources into the System of Social Statistical Datasets (SSD), Statistics Netherlands faces these limitations.

134. Everyone who enters the Netherlands is registered as a resident (immigrant) provided that:

- His/her stay is legal according to the Immigration Act (for people who do not have the Dutch nationality);
- The intended stay is at least two thirds of the forthcoming six months;
- The person is properly identified. The latter means that a valid passport or other official document is shown for identification.

135. Every child born in the Netherlands whose mother is registered as a resident is also registered as a resident. Children who are born abroad to a mother who herself is registered as a resident of the Netherlands are registered (as immigrants), provided that the children will live in the country.

136. Emigration relates to persons who leave the Netherlands and intend to stay abroad for at least two thirds of the forthcoming twelve months, and who inform the municipal authorities of their departure. They are included in the Dutch emigration statistics.

4.1.6.2 Limitations of the original or main source. Why is integration necessary?

137. There are two exceptions to the rules for registration of immigrants, which make the registrations not entirely complete.

138. The first applies to so-called 'privileged persons', including foreigners working on Dutch soil as diplomatic, consular or military officials, or in an international organization. Since they benefit of a special 'privileged' status and are not considered foreigners under the auspices of the Immigration Act, they are given the choice of whether or not to be entered in the population register.

139. Asylum seekers constitute another exception. Their registration takes place only six months after their arrival in the Netherlands, irrespective of their expected duration of stay, unless they are granted a residence permit within six months. In that case they are registered when the residence permit is granted. However, children born to asylum seekers who are not yet registered are registered directly after birth. This leads to the somewhat odd situation that a newborn baby is registered whereas the parents and siblings are not (yet).

140. When it comes to emigration, about one in three persons who leave the country does not notify the municipal authorities of his or her departure. When the authorities find out that someone is 'missing', the law stipulates that they must investigate his or her whereabouts. If the investigations lead to the conclusion that the person remains missing, he or she is administratively registered as emigrated to a country qualified as 'unknown'. They are included in the statistics of administrative removals. If they return to the Netherlands they are included in the statistics of administrative entrances.

4.1.6.3 Description of the methods used for data integration

141. All production processes within Statistics Netherlands concerning social or spatial statistics are based on the System of Social Statistical Databases (SSD). The SSD contains a wealth of information on persons, households, jobs, benefits, pensions, education, hospitalizations, crime reports, dwellings, vehicles and more. Before the SSD, various statistical registers were scattered within Statistics Netherlands and were not standardized. By the process of micro-integration, the SSD resulted in a library of standardized and linked statistical registers, as well as the measures which have been put in place to control various aspects of the system.¹⁵

142. When new registers come to Statistics Netherlands, they also converge into the SSD, using the same system of standardization, linkage and micro-integration. Recently, Statistics Netherlands added data from the Central Agency for the Reception of Asylum Seekers (COA) to the SSD. This will solve the problem as described above, since asylum seekers will soon be registered from the first day on which they enter the Netherlands. Hopefully other sources, such as PROBAS (a database containing information on privileged persons) or RNI (a municipal registration of non-residents) will soon be added to the SSD.

¹⁵ For more information on the SSD see: Bakker, B., Van Rooijen, J., & Van Toor, L. (2014). The system of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics. *Journal of the International Association for Official Statistics*, 30: 411–424.

4.1.6.4 Benefits of integration with statistics from other sources

143. The benefits of integration with statistics from other sources are evident. First of all, integral data is available, reducing the need for surveys. This reduces the costs of gathering data and increases the number of respondents. The benefits of standardization and micro-integration of data are that information from different sources can be linked to the same persons. This results in a data library containing more than fifty administrative registers. The integrated data of Statistics Netherlands therefore offer a rich source of data, which not only allows researchers to measure migration, but also gives the opportunity to study socio-economic integration with a broad range of variables.

4.1.7 Spain¹⁶

4.1.7.1 Introduction

144. INE-Spain does not currently apply data integration techniques to produce migration statistics. The only source from which these statistics are obtained is the population register, although statistical treatments are performed on the raw data included in the population register to improve the aggregate figures on flows.

145. However, the 2011 population census in Spain was a clear example of integration of different sources to improve the measurement of population stocks, and especially the stock of foreign residents in Spain. The migratory flows for the period 2008-2011 were consequently revised. The method used for data integration in the 2011 census is described below.

4.1.7.2 Primary source of data for migration statistics: the population register ('Padrón')

146. The main source of data in Spain for both population stocks and migration statistics is the population register, named Padrón in Spanish. Padrón is the official list of residents in each of the 8,100 municipalities in Spain. Padrón, as a list of residents of every municipality, comes from a very old tradition in Spain, dating from the Middle Ages.

147. There are as many registers as there are municipalities in Spain. But there is a law, in force since 1996, integrating these municipal lists into a single national database. There are also legal procedures to keep this database and the municipal files interconnected and updated on a monthly basis.

148. The law requires municipalities to exchange these records. This is done through the statistical office of Spain, INE. So, unlike other countries where the police or other administrative bodies are in charge of population registers, in the case of Spain, INE is the national institution that coordinates this single national population register. Every month INE receives information about all changes that have taken place in every municipality, performs validations and forwards these results to all the municipalities, to avoid duplications and also to include deaths, births or acquisition of Spanish citizenship that INE receives on a monthly basis (or more frequently) from the Civil Register.

149. All consular offices of Spain throughout the world (around 250) are also connected to Padrón in the same way as the municipalities. Any Spanish national leaving a municipality within Spain to live abroad should declare his/her new residence at the closest consular office. Consular offices send this information to INE and INE forwards these data to the municipalities in the monthly coordination files.

150. According to the law, there is no restriction, in terms of legal situation, for registering in Spain. All people living or wishing to live in Spain, regardless of their legal situation have the right to be registered (in fact this is an obligation) and they normally are, since being registered brings many advantages and no drawbacks. The benefits of being registered may comprise free access to the public health system, public schools, services and rights granted by the municipalities, as well as cheaper transport and many others.

151. For each person, the register contains sex, date of birth, place of birth (country, in the case of foreigners), nationality, educational attainment and the national identity number. Legally resident foreigners have an ID number. For people with an irregular status the passport number is stored.

152. Padrón is also a longitudinal database: all previous places of residence (within Spain) are also stored, thus allowing longitudinal analysis and a very close and precise monitoring of internal migration.

¹⁶ Prepared by the Directorate for Socio-Demographic Statistics, INE.

4.1.7.3 Limitations of the original source. Why is integration necessary?

153. As noted, Padrón is the source for population figures and migration statistics. This does not mean that the population count directly corresponds to the number of people registered at a given moment. Some statistical treatments are applied to the raw data included in the population register to provide the population figures.

154. The main drawback of a system based on a population register like this one is the difficulty in measuring emigration. With regard to immigration, registered inflows are likely closer to the true values. People may find many reasons to register but not many for de-registering.

155. When the population register was set up in 1996, the number of foreign nationals was 540,000 (1.4 per cent of the population of Spain - 39.7 million at that time). Some years earlier, in the census of 1991, it was almost the same (470,000). Spain had a very small net international migration in the first half of the 1990s.

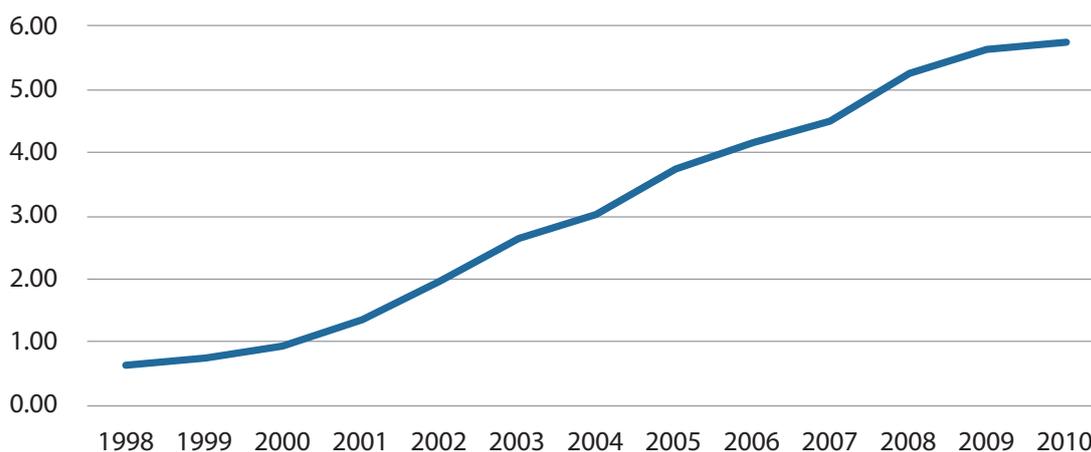
156. The second half of the 1990s, and especially the first decade of 21st century were very different. Because of the growth of the economy, the number of foreign residents increased dramatically. In only ten years (2000 to 2010) the number of foreign residents according to the population register grew from 1 million to nearly 6 million (from 1.4 per cent to 12.2 per cent of the population in 2010).

157. The register had to address an important challenge for which it had not been designed: how to count foreign people, and more specifically how to measure emigration. No process had been designed for de-registering foreign people when they did not declare their departure.

158. A new legal procedure was approved in 2006. According to their legal status, foreign residents can be divided into two groups: EU citizens or non-EU but having a permanent residence permit (let us call them 'EUP' foreigners) and the rest, i.e., non-EU citizens that hold a non-permanent residence permit or do not hold a residence permit ('non-EUP' foreigners).

159. Since 2006, all non-EUP foreigners (roughly 50 per cent of total foreigners at that time) are required to renew their registration in Padrón every two years. Every city council must send notifications to all non-EUP foreigners whose registrations are about to expire. If they do not renew their registration, after an administrative process that may take some months, they are de-registered from the municipal records.

Figure 4.3 Foreign population in Spain according to the population register 1998-2010 (millions)



Source: Population Register (Padron). INE

160. A similar process was set up for EUP citizens but it was built later, in 2009, and was not fully applied until 2016.

161. At the time when the decisions on the population census had to be taken, around 2008-2009, there was considerable uncertainty about the real number of foreign residents in Spain. How many of the nearly 6 million foreigners registered since 1996 had left the country without declaring so to the public administration? How many Spaniards had left the country without declaring a new residence abroad to consular offices?

162. Eventually, in 2009, the methodology for the 2011 census was outlined. It was conceived as a combination between registers (linking Padrón with others) and a survey addressed to 10 per cent of population, mainly used to collect all demographic variables not included in registers.

163. The main concern when designing the census methodology was how to combine registers to improve the measurement of population stocks, both for Spaniards and foreigners, considering that there was little evidence about outflows in the population register, notably for the foreign population. In the case of Spaniards, at least there is a source of information on outflows, namely the declaration at consular offices.

4.1.7.4 Methods used for data integration

164. The 2011 population census included a complex procedure of data integration to improve the population count. This procedure was used to count the so-called 'doubtful' population, i.e. people with registered residence in Padrón without evidence of residence in Spain in other administrative registers (tax collection agency and social security databases, among others).

165. During the 2011 census, in order to count the population stock, the registered population (both Spaniards and foreign residents) was linked with many different administrative registers to provide to each person a likelihood of residence, a number called the 'count factor'. As a result of this process a 'pre-census' file was built. The main principle is simple. If a person is correctly registered and if he/she is receiving a salary or a pension, or paying taxes during 2011, then this person is very likely to be residing in Spain. These verified records are named 'sure' population. But people not found in other records are considered 'doubtful' population.

166. The pre-census file was the result of this process, with roughly 47.4 million records.

167. Around 15 different criteria by which a record could be considered doubtful were defined. After this linkage, the results were:

- 97.7 per cent of the pre-census file records (46,372,000) were considered 'sure' population (a count factor equal to 1 was assigned to each person in this group);
- Around 40,000 records (0.1 per cent of the population) were removed from the pre-census file after a probabilistic record linkage with death records. They were considered errors in the population register and were assigned a count factor equal to 0;
- For the remaining 2.2 per cent of records the evidence was inconclusive, so they were considered doubtful records. There were 1,046,000 such records, and 87 per cent of these were foreigners. Therefore, their 'count factor' was unknown.

168. The first finding of this process was that the doubtful population consisted mainly of migrants, but not economic migrants. During the decade before the 2011 census, Spain had received large flows of migrants from Ecuador, Morocco or Romania, but there were also significant and continuous inflows of Germans or Britons who came to Spain to live after their retirement. These 'good weather migrants', often referred to as sun-seekers, are more difficult to find in registers: they can be registered in Padrón but they do not receive a pension from the Spanish government nor do they pay taxes in Spain.

169. The 'count factor' for doubtful records was calculated using the 'census survey' (as mentioned, the 2011 census in Spain comprised a survey addressed to 10 per cent of population).

170. For this purpose, doubtful records were grouped into population classes defined by socio-demographic characteristics: age groups, nationality and province of residence (there are 52 provinces in Spain). The pre-census file was partitioned into these classes. Each class contained sure and doubtful records. If a class did not contain any doubtful record, then no estimation was needed, and the population for that class was equal to that from the pre-census file.

171. All classes containing doubtful records were configured to include at least 1,000 such records. Classes with less than 1,000 doubtful records were grouped in order to reach that threshold. The grouping strategy was first to select wider age groups, then wider territory areas and lastly grouping nationalities. 724 clusters or classes were created, with at least 1,000 doubtful records in each.

Table 4.2 Doubtful records in the 2011 pre-census file. Selected nationalities

Nationality	Total Population in pre-census file with CF* ≠ 0	Doubtful records	% Doubtful
Total	47,418,916	1,046,433	2.2
Spain	41,589,484	136,452	0.3
Romania	886,631	120,216	13.6
United Kingdom	394,998	112,100	28.4
Morocco	806,012	61,459	7.6
Germany	195,667	58,690	30.0
Bolivia	197,113	37,492	19.0
Paraguay	90,788	32,107	35.4
Bulgaria	173,061	31,269	18.1
Brazil	105,216	28,352	26.9
France	120,780	27,482	22.8

*CF= count factor

172. In order to calculate these count factors we needed to estimate the proportion of sure population in the survey. In a given class i of the pre-census file, the total number of records, T_i was split into the number of records of sure population S_i and the doubtful ones D_i ,

$$T_i = S_i + D_i$$

173. To improve the estimation of the total population, the count factor was applied to doubtful records in class i as shown below to create a new estimation of total population for this class, T_i^* :

$$T_i^* = S_i + CF_i * D_i$$

174. The population from the survey was grouped into the same 724 classes according to the same characteristics (age, territory, nationality). After that, the sample and the pre-census file records were linked at the individual level to determine which records from the sample had been previously classified as doubtful in the pre-census file.

175. With those count factors incorporated into every doubtful record, the pre-census file became the Weighted Census File (WCF) and allowed for the derivation of census population figures. The population of a given geographic area, T_i^* , was obtained as the sum of the count factors of the WCF records in that area (given that CF=1 for all sure records).

176. This procedure provided statistical figures for the census, but of course, it cannot be used for administrative purposes to update the population register, since it is not possible to determine which individual citizens should be counted and which should not.

177. The most noteworthy class, in terms of the proportion of doubtful population, was that of German nationals aged 80 and over in the province of Alicante (a place on the Mediterranean coast), comprising nearly 3,000 persons, 70 per cent of whom were considered doubtful. The largest class in terms of absolute number of doubtful persons is that of citizens of the United Kingdom, also living in the province of Alicante, aged 60 to 65 years old. Comprising 5,963 doubtful records, out of 19,317 UK citizens included in this group in the pre-census file.

178. Some selected cases are shown in Table 4.3.

4.1.7.5 Benefit of integration with statistics from other sources

179. As stated before, the number of doubtful records was 1,046,000 in round numbers. The average count factor for all classes was 0.424, which means that these doubtful records were counted as a population of approximately 440,000 people.

180. The CF was higher than 1 in only 22 classes, meaning that these populations were considered to be under-registered. This was the case for nationals of Pakistan aged between 25 and 29, with a count factor of 1.37 and affecting 1,055 doubtful records (subsequently counted as a population of 1,445).

Table 4.3 Doubtful population in the pre-census file group by nationality-province-age

Group (nationality, province, age)	Population in pre-census file (T_i)	Sure records (S_i)	Doubtful records (D_i)	% doubtful records	Count factor (CF_i)	Estimated population $T_i^* = S_i + (D_i * CF_i)$
United Kingdom, Alicante, 60 to 64 years	19,317	13,354	5,963	30.9%	0.347	15,426.1
United Kingdom, Alicante, 65 to 69 years	21,180	15,287	5,893	27.8%	0.222	16,598.1
Germany, Alicante, 70 to 74 years	6,703	2,633	4,070	60.7%	0.269	3,729.1
United Kingdom, Alicante, 55 to 59 years	11,732	7,663	4,069	34.7%	0.254	8,695.3
Romania, Madrid, 25 to 29 years	32,909	29,867	3,042	9.2%	0.257	30,649.8
Pakistan, Spain, 25 to 29 years	4,384	3,329	1,055	24.1%	1.370	4,774.5
Morocco, Madrid, 0-9 years	17,162	16,158	1,004	5.9%	0.115	17,046.6

181. As a result of these calculations, the population of Spain reached 46,815,916 inhabitants as of 1 November 2011, some 450,000 inhabitants below the figure provided by the population register at the same date. While the number of Spaniards hardly changed, the stock of foreigners according to the census decreased by 480,000 compared to the population register figures to reach a total number of 5,252,473 (11.2 per cent of the population of Spain, 8 per cent fewer foreigners than the figure given by the population register).

182. The most prominent case of decrease in population figures was that of UK nationals. While according to the register there were 391,194 UK nationals, the number according to the census was only 312,098.

183. In conclusion, the use of data integration techniques, combining administrative data and survey-based data, improved the population figures, mainly on foreign nationals. Moreover, since 2011, the population register has been improved greatly. Procedures have been put in place in town councils to verify the effective residence of foreign population, especially 'EUP' foreigners, and to improve the population figures provided by the population register.

4.1.7.6 References

http://www.ine.es/en/censos2011_datos/cen11_datos_metodologia_en.htm

Argüeso, A. and Vega, Jorge L. (2013) "A population census based on registers and a "10% survey" methodological challenges and conclusions". Statistical Journal of the IAOS – Vol. 30

4.1.8 Switzerland

184. The Swiss Federal Statistical office (FSO) does not currently apply any data integration techniques in the strictest sense to produce migration statistics. The only sources for such statistics are the population registers.

185. The production of annual migration data is an integral part of the Population and Households Statistics (STATPOP). It is based on one of the various surveys conducted within the framework of the federal population census. The statistics provide information regarding population size and composition at the end of a given year as well as population change and its demographic components during the same year.

186. In 2010 the traditional decennial census was replaced by an integrated statistical system which provides annual data. The new census system combines the use of administrative registers with sample surveys. It consists of four different annual surveys, among them – and of particular importance to migration statistics – the register survey.

187. The register survey is the source for STATPOP. It exploits existing and harmonized administrative data from both centralized and local population registers, and is therefore particularly well suited for producing data on migration flows.

Registers are dependent on the timely registration of the population and keep track of all changes of residence, thus ensuring a continuous update of individual records. Administrative sources are frequently able to do so, despite some of their well-known drawbacks (e.g. statistical data collection is not a priority, definitions and coverage depend on legislation and administrative rules).

188. The nationwide social security number which uniquely identifies a person (universal PIN) plays an important role both in the maintenance of administrative registers and in their statistical use, particularly in linking information from register and survey data.

4.1.8.1 Primary data sources for the measurement of migration stocks and flows: the population registers

189. The main data sources for official annual migration statistics (demographic stocks and flows) are the administrative registers at the federal, cantonal and municipal levels.

190. There are three different types of registers:

- a) 2,300 local population registers (maintained by the municipalities or in some cases the cantons¹⁷) for the national and non-national population;
- b) Central Migration Information System (an aliens register maintained by the State Secretariat for Migration at the federal level) for the legally resident non-national population (i.e. people in possession of an official permit of stay);
- c) 'Ordipro' information system (an aliens register maintained by the Federal Department of Foreign Affairs) for non-nationals that are entitled to privileges and immunities and not subject to Swiss immigration laws, i.e. staff of diplomatic missions, consular posts, permanent missions and intergovernmental organizations (including spouses, partners, unmarried children as well as private household employees).

191. Whereas Swiss citizens are registered at the local level only, foreign nationals can appear in several registers (in one or more local population register as well as in one of the two federal registers). Since there is no single data source which provides a comprehensive picture of migration of non-nationals, it is good practice to use multiple sources. While in some instances these sources will lead to the same findings, in other instances the details and trends may differ. However, this does not necessarily mean that one source is 'right' and another is 'wrong' or that one source is 'better' than another one.

192. Therefore, for all categories of foreign nationals (e.g. permanent residents, asylum seekers etc.) it is necessary to select the source on which the number of migrants (stocks) or migratory flows to be considered in official statistics is to be based. Timeliness, completeness, reliability and other factors are critical to prioritize the sources.

193. The following decisions have been established for the production of annual migration stock and flow data:

- The register of reference for the 'permanent resident foreign population' (long-term migrants, i.e. stocks of persons who have resided in Switzerland for at least 12 months and flows of persons who arrive in and leave Switzerland and have been issued legal settlement or residence permits that are valid for at least 12 months) are the local population registers;
- The register of reference for the 'non-permanent resident foreign population' (short-term migrants, i.e. stocks of persons who have resided in Switzerland for less than 12 months and flows of persons who arrive in and leave Switzerland and are in the possession of legal permits that are valid for less than 12 months) is the Central Migration Information System;
- The register of reference for foreign nationals who have been issued legitimation cards by the Federal Department of Foreign Affairs is the Ordipro information system.

¹⁷ There is no central population register covering the entire (national and non-national) population of Switzerland. In addition, FSO does not maintain a statistical population register. It receives data extracts containing pre-defined variables in a pre-defined standardized data format from each register at regular intervals.

194. Consequently, 'data integration for measuring migration' in Switzerland refers to the deliberate choice of one administrative register rather than another for specific sub-groups of the non-national population.

195. The local and the federal data sources are not mutually exclusive. They have partially overlapping observation units as well as partially overlapping variables. This overlap is due to the fact that the local population registers are supposed to include all resident foreign nationals, who in addition should be included in one of the two federal registers. However, no attempt is made to systematically compare the registers in order to identify individuals (stocks) and migratory events (flows) that are recorded in multiple registers or in a single register only (and therefore missing elsewhere). The count of migrants and migration inflows and outflows is neither augmented by observations found in an administrative database that is not considered to be the 'register of reference' for a specific population sub-group, nor are observations in the 'register of reference' disregarded or deleted if they are not found in another register.

196. Moreover, the characteristics relating to an individual can differ from one register to another. Depending on the task or the position of the register in the administrative process, from a statistical point of view it can be assumed that the quality of some of the characteristics or variables (such as sex, age, marital status, citizenship, country of last/previous residence etc.) is better or more coherent in some registers than in others. Therefore it may be necessary to prioritize the sources for the modalities of certain characteristics. For the production of annual migration statistics it is again the 'register of reference' which determines which modality is taken into account.

197. However, the Central Migration Information System provides some exclusive variables that are missing at the local level. They are added to the local population register records pertaining to the same observations (individuals or migratory events). These variables include 'nationality of spouse/partner', 'purpose of stay' or 'migration motive', 'date of first issue of permit', 'expiry date of permit' and, for asylum seekers, detailed information about the various administrative steps of the asylum procedure.

198. This data 'enhancement' procedure has been implemented into the annual migration data production and can be considered a modest form of data integration. Data collected at the federal level are matched with data collected at the cantonal or municipal level. The availability of the social security number in each of the involved registers allows the unambiguous linking of records of one individual in different sources.

199. In the field of migration statistics, FSO makes a clear distinction between statistics and estimates, the former broadly representing the product of an (as far as possible) exhaustive compilation of records from primary data source(s), the latter the outcome of probabilistic/statistical models which might involve combining information from various sources. Official Swiss migration statistics do not include any estimations – neither at the individual nor at the aggregated level.

4.1.8.2 Other data sources for the measurement of migration stocks and flows not integrated into the production of annual migration statistics

200. There are additional sources of migration data in Switzerland, the two most important ones being the Swiss Labour Force Survey (LFS) and the Structural Survey (an annual sample survey of 200,000 people which is part of the new census system and which collects variables not currently available in the registers, e.g. language, religion, educational attainment).

201. A key aspect of the new Swiss census system is the possibility of linking of data from population registers with data from an annual sample survey by means of the social security number. The information contents of both sources can thus be combined to supplement each other. However, at present no such matching or combining of data is being done within the context of annual migration statistics.

202. Migration-relevant information from sample surveys are presented separately and provide complementary insights into stocks and flows of migrants. Moreover, survey data and register data do not only differ in geographic scope and data collection methods but also in the population universe. Both the LFS and the Structural Survey cover only persons aged 15 and over living in private households.

4.2 Case studies of countries without population registers

4.2.1 Australia

4.2.1.1 Sources of migration statistics

203. Discrepancies Administrative information on persons arriving in, or departing from, Australia is collected via various processing systems, passport documents, visa information, and incoming and outgoing passenger cards (see Appendix 1 of reference below). Incoming persons provide information in visa applications except those travelling as Australian or New Zealand citizens. These administrative data are collected by the Australian Government Department of Immigration and Border Protection (DIBP) under the authority of the Migration Regulations (Migration Act, 1958).

204. The Australian Bureau of Statistics (ABS) statistics on overseas arrivals and departures (OAD) are compiled using information from DIBP sources. Overseas movements are collected and matched (where possible) by DIBP and then stored with movement records on the Travel and Immigration Processing System (TRIPS). Each month the matched OAD movement records are supplied to the ABS and then processed. A unique personal identifier is the mechanism used for linking all the data sets at DIBP and then used by the Australian Bureau of Statistics to generate traveller histories of each individual traveller.

205. Quarterly net overseas migration (NOM) estimates are sourced from this processed monthly OAD matched data, and then combined with monthly extracts of unmatched OAD records. Unmatched OAD records are those where an inward or outward movement has been recorded by DIBP within the TRIPS system, but it has not been possible to match the data with an equivalent passenger card. Statistics on overseas migration exclude: multiple movements; the movements of operational air and ships' crew; transit passengers who pass through Australia but are not cleared for entry; passengers on pleasure cruises commencing and finishing in Australia, and undocumented arrivals. From 1 July 2006 onwards, foreign diplomatic personnel and their families are also excluded.

206. For a reference to the methods on how Australia measures international migration see <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3412.0Explanatory%20Notes12016-17?OpenDocument>

4.2.1.2 Sources for measuring migration stock

207. As mentioned above, for international migrants the source for country of birth is data provided from Australia's DIBP which is a combination of passport and visa information. For deaths, the source for country of birth is information provided by the Birth & Death registrars in each State and Territory of Australia. All births are, of course, recorded as Australian-born. In addition, the Australian Census is used once every five years to re-base the Australian Population including by country of birth.

4.2.1.3 Use of data integration with other sources

208. One of the most useful variables has been a unique personal identifier (PID) for each individual who crosses the Australian border, whether in or out. This allows many additional uses for the data including linking an individual over time and within various systems.

209. The Australian Bureau of Statistics, in collaboration with the Australian Department of Immigration and Border Protection, links migrant data sets with census data. In addition, recent work has also linked this same data with tax information - see relevant document below.

210. See the following releases by most recent first:

Migrant Data Matrices, 2017 ¹⁸

Understanding Migrant Outcomes - Insights from the Australian Census and Migrants Integrated Dataset, Australia, 2016 ¹⁹

Microdata: Australian Census and Migrants Integrated Dataset, 2016²⁰

¹⁸ <http://www.abs.gov.au/AUSSTATS/abs@.nsf/productsbyCatalogue/70AA0E84BE9D586ACA2575400017B0F3?OpenDocument>

¹⁹ <http://www.abs.gov.au/AUSSTATS/abs@.nsf/productsbyCatalogue/706907E56F9F5128CA257BEA00111584?OpenDocument>

²⁰ <http://www.abs.gov.au/AUSSTATS/abs@.nsf/productsbyCatalogue/C916B16440BF9B60CA257C7E000FC851?OpenDocument>

Perspectives on Migrants, Mar 2013 ²¹

See also the Guide to all the other Migrant Statistical Sources ²²

4.2.2 Canada

4.2.2.1 Original/primary source of data for migration statistics

4.2.2.1.1 For population estimation purposes

211. Immigration, Refugees and Citizenship Canada (IRCC) collects and processes immigrants' and non-permanent residents (NPRs) administrative files. It then provides Statistics Canada (STC) with the information.

212. For immigration, the files are used to estimate the number and characteristics of people granted permanent resident status by the federal government for a given period.

213. For NPRs, the files required to produce their estimates include visitors' permits, work permits, study permits, Ministers' permits, refugee status claims, landings, applications for landing and deportations. The information is used to estimate the number and characteristics of people granted non-permanent resident status by the federal government.

214. Emigration estimates require a distinction between persons establishing a permanent residence in another country (i.e., emigrants), persons living temporarily abroad while not maintaining a usual place of residence in Canada; others who temporarily leave Canada and then return (i.e., net temporary emigrants), and finally the portion of emigrants who have returned to Canada (i.e., returning emigrants).

215. Emigrants are estimated from administrative sources in terms of the gross flow of migrants out of Canada. The Office of Immigration Statistics of U.S. Department of Homeland Security provides data on Canadians who acquire permanent immigrant status in the U.S. This data source is used for estimating emigration to the United States. In order to estimate emigration to other countries, information on notification of departure from the Canada Child Tax Benefit (CCTB) programme and tax data from the Canada Revenue Agency (CRA) is used. Various adjustments are made to the data, for example to correct for incomplete coverage.

216. Net temporary emigrants are estimated with several sources. Data from the Reverse Record Check, the census coverage study used to measure undercoverage, are used to estimate the number of persons leaving the country temporarily; while data from the National Household Survey (NHS), combined with Demography Division's estimates of returning emigrants, are used to estimate the number of temporary emigrants returning.

217. Returning emigrants are Canadian citizens or immigrants having previously emigrated from Canada and subsequently returned to Canada to re-establish a permanent residence. Again, data from the CCTB programme and tax data are used for estimating returning emigrants.

4.2.2.1.2 For analytical characteristics

218. The permanent and non-permanent resident data provided by IRCC offers relevant information for analytical research and policy needs. This data source is a census of the immigrant population and, as such, offers the capacity to study small populations in detail. The data source relating to permanent residents contains socio-demographic information on immigrants who have landed since 1980. Characteristics related to knowledge of languages, destination in Canada, country of origin, intended occupation or level of education are obtained at the time of admission.

219. For NPRs, pre-admission files provide detailed geographic information and the main activity of the individual such as work or study during specific periods after 1980.

220. The Census of Population provides a statistical portrait of the country every five years. The 2016 census identifies permanent and non-permanent residents living in Canada. It provides many socio-economic characteristics of immigrants such as family composition, knowledge of languages, labour force status and income.

²¹ <http://www.abs.gov.au/AUSSTATS/abs@.nsf/productsbyCatalogue/3B2787D4377D2D84CA2573F7000DDE5B?OpenDocument>

²² [http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/3414.0Main%20Features42011%20\(Edition%202\)?opendocument&tabname=Summary&prodno=3414.0&issue=2011%20\(Edition%202\)&num=&view=](http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/3414.0Main%20Features42011%20(Edition%202)?opendocument&tabname=Summary&prodno=3414.0&issue=2011%20(Edition%202)&num=&view=)

4.2.2.2 Limitations of the original or main source. Why is integration necessary?

4.2.2.2.1 For population estimation purposes

221. Immigration data are straightforward administrative data. However, for provincial and territorial estimates, the file obtained from IRCC indicates the province or territory of intended destination on arrival, rather than the province or territory in which the immigrant actually settles. In a small number of cases, information on the province of destination is missing. These cases are distributed proportionally between the provinces and territories according to the observed distribution of immigrants for whom this information is available.

222. For data on NPRs, there is the same limitation relating to the intended province of settlement as for immigration data. There is also a lack of information on the departure dates of NPRs, as some of them leave the country before the end of the validity of their permits. Since it is not mandatory to inform the federal government before leaving the country, STC does not have any additional sources to obtain this information.

223. Moreover, both immigrants and NPRs are assigned a Client Identification Number (CID) by IRCC. This identification number is a key variable as it allows identification of each person only once. The CID is used in the production of population estimates because it allows STC to create a biography of each individual, even if the person holds more than one permit at the same period.

224. Unlike immigration, there is no legal provision in Canada to maintain records for persons leaving the country either on a temporary or permanent basis. Therefore, estimates of the number of emigrants and persons living temporarily abroad and their characteristics must be derived through secondary sources such as Canadian administrative files or immigration statistics of the U.S. Also, few data sources allow for a measure of emigration, and these sources often use different definitions of emigration. Given its nature, temporary emigration is notably hard to measure. As a result, various evaluations suggest that Canada underestimates the levels of emigration and temporary emigration. In this context, integration is necessary to make the best possible use of the strengths of the various data sets. For instance, since the U.S. is the country of destination of most emigrants from Canada, using American immigration data is thought to improve our measure of Canadians moving to the U.S.

4.2.2.2.2 For analytical characteristics

225. The administrative data from IRCC does not provide information on any outcomes after admission. Information related to socio-economic outcomes, changes in the family structure, and changes in residence may only be obtained from linkage to external data sources such as annual tax files.

226. The Census of Population does provide up-to-date outcomes on immigrants after landing, but it does not include any information on admission conditions or selection criteria.

4.2.2.3 Methods used for data integration

4.2.2.3.1 For population estimation purposes

227. In the case of immigration and NPRs, information is integrated at the level of individual records. All the files contain the same variables and are linked using a unique CID. Measuring the number of immigrants entering Canada in a given period is straightforward, and adjustments to the data are not required. Information is available for each person entering Canada under immigrant status from IRCC's administrative file. Every month, IRCC makes available to STC a data file containing the records of immigrants for the previous month, as well as any additions or updates to data already received.

228. For the NPR population, there are two major subgroups that are administratively different (permit holders (PH) and refugee status claimants (RSC)); their estimates must, therefore, be produced separately. For more information regarding the methodology of the RNP, please refer to <http://www.staging.statcan.gc.ca/pub/91-528-x/2015001/ch/ch5-eng.htm#n1>

229. As for emigration (emigration, net temporary emigrants and returning emigrants), please refer to the methodological document: <http://www.staging.statcan.gc.ca/pub/91-528-x/2015001/ch/ch6-eng.htm>

4.2.2.3.2 For analytical characteristics

230. To circumvent the limitations of the administrative immigration data, Statistics Canada uses exact matching record linkage techniques to combine the information with other data sources such as tax files. The Longitudinal Immigration Database (IMDB) is updated every year and includes all immigrants who have landed in Canada since 1980 with outcomes from annual tax files since 1982. For processing year 2014, 89 per cent of immigrants were linked to the tax files. As such, the database is an administrative census with a longitudinal design, and therefore no sampling is done.

231. In addition, the 2011 National Household Survey was linked with immigration administrative data to connect admission conditions with the socio-economic outcomes available in the survey. For the 2016 Census of Population, record linkage was used to add variables related to admission category to the census database.

4.2.2.4 Benefit of integration with statistics from other sources

4.2.2.4.1 For population estimation purposes

232. The main benefit of integrating different sources of data is to provide estimates of greater quality that will be useful not only for STC, but also for its many external users. Finally, STC is currently working on a Statistical Population Register. It would then be possible to link immigration and emigration data through a unique personal identifier.

4.2.2.4.2 For analytical characteristics

233. The record linkage between the Census survey data and the immigration administrative file provides additional analytical capacity while reducing the burden on respondents. This approach is a good alternative to obtain details on immigration selection category that respondents may neglect, overlook, or cease to remember over time.

234. Being mostly transactional in nature, IRCC immigration administrative files are somewhat restricted with regard to performance and research indicators. The IMDB is created from the record linkage between immigration and tax data files to provide a more comprehensive source of data on socio-economic outcomes of immigrant tax filers. This data source contains detailed information on labour market behaviour and covers a period long enough to assess the impact of characteristics at admission, such as education or knowledge of French or English. Elements such as pre-admission experience of immigrants (i.e. study experience in Canada), mobility within Canada, or changes in the family composition take advantage of the longitudinal aspect possible with each new tax report.

4.2.3 New Zealand

235. As the principal agency responsible for the processing of international movements into and out of New Zealand, Statistics New Zealand provides timely and accurate statistics on three passenger types: overseas visitor, New Zealand-resident travellers, and permanent and long-term migrants. Having an advantageous geographical situation with border entry by sea ports and airports only, travellers' self-completed passenger cards have been the main form of data collection.

236. The content of the cards has evolved over time, and in the longer term there are expectations to reduce further the information passengers provide on the paper forms. There has also been a strong reliance on integrated electronic files of international passenger movements as a source for validation. Statistics New Zealand is investigating the use of alternative sources and statistical methods for obtaining the departure information, with a view to reducing or removing the reliance on the departure card.

4.2.3.1 Data collections

4.2.3.1.1 Sources for measuring migration flows

237. Releases of New Zealand international migration statistics are based on the electronic capture of all passenger arrivals and departures supplied to Statistics New Zealand by the New Zealand Customs Service. The records include travel identities such as carrier details, travel mode and port, as well as key passport information such as date of birth, sex and country of citizenship. Records include the processing of passengers who require a visa to enter New Zealand. Visa types are categorized according to the policy and immigration decision-making requirements of Immigration New Zealand. All border movements and passenger data captured by the New Zealand Customs Service are regularly (daily) transferred to Statistics New Zealand.

238. The New Zealand Customs Service also supplies Statistics New Zealand with the physical passenger cards which were completed by travellers at the time of travel. Statistical information captured by the passenger card includes additional traveller information such as country of last or next residence, travel purpose, and intended length of stay in New Zealand or absence from New Zealand. It also enables the classification of passenger categories (long-term, overseas visitor, resident).

239. Statistics New Zealand undertakes the scanning of the passenger cards using image recognition technology for automatic coding of text responses. This is followed by a manual completion process of records not clearly recognized and coded by the imaging software (about one in 10 cards). Some response fields are required from each passenger card for the purpose of linking to the electronic records supplied by New Zealand Customs Service. Other response fields of the passenger card are only required for sampling purposes and vary by passenger type and direction of movement

4.2.3.1.2 Sources for measuring migration stocks

240. The five-yearly population census provides measures of the overseas-born population by capturing the country of birth and the year of arrival in New Zealand for the usually resident population. At the 2013 census, the overseas-born resident population count was around 1 million, representing nearly one quarter of the total New Zealand usually resident population (Figure 4.4). In 1986, the overseas-born population of New Zealand was less than 15 per cent and there have been marked increases in the number and proportion of the resident overseas-born population until 2006 (Figure 4.4 and Figure 4.5).

241. A post-enumeration survey (PES) is undertaken shortly after the census to evaluate the completeness of census coverage. Estimates of under-enumeration and over-enumeration are produced by the PES for population groupings defined by age, sex, ethnic groups and geographic area. The five-yearly historic PES (1996-2013) have not provided estimates of the net census undercount of the overseas-born population (Statistics New Zealand, 2014). However, for the 2018 PES, Statistics New Zealand has made plans to include the country of birth question. This is acknowledging the need to improve accuracy and to extend statistics of the overseas-born population.

242. Table 4.4 summarizes the data content of the migration data collections combined for the production of monthly statistics on international passenger movements.

243. A passenger may change their intentions of length of stay or absence after their arrival or departure, which may result in the recorded passenger type at time of travel becoming incorrect. Statistics New Zealand does not revise published statistics to adjust for such changes. There is a need for an alternative method for classifying passengers after having observed their actual patterns of stay in New Zealand. A time series of linked travel sequences for individual travellers enables a final classification of their migrant status.

4.2.3.1.3 Sources for measuring migration stocks

244. The five-yearly population census provides measures of the overseas-born population by capturing the country of birth and the year of arrival in New Zealand for the usually resident population. At the 2013 census, the overseas-born resident population count was around 1 million, representing nearly one quarter of the total New Zealand usually resident population (Figure 4.4). In 1986, the overseas-born population of New Zealand was less than 15 per cent and there have been marked increases in the number and proportion of the resident overseas-born population until 2006 (Figure 4.4 and Figure 4.5).

245. A post-enumeration survey (PES) is undertaken shortly after the census to evaluate the completeness of census coverage. Estimates of under-enumeration and over-enumeration are produced by the PES for population groupings defined by age, sex, ethnic groups and geographic area. The five-yearly historic PES (1996-2013) have not provided estimates of the net census undercount of the overseas-born population (Statistics New Zealand, 2014). However, for the 2018 PES, Statistics New Zealand has made plans to include the country of birth question. This is acknowledging the need to improve accuracy and to extend statistics of the overseas-born population.

Table 4.4 International migration data collections and processing

Data collections		
Movement records	Passenger records	Imaged passenger cards
Actual date	Customs port	Travel date
Scheduled date	Passport number	Passport number
Prime carrier/ ship name	Citizenship	Carrier/ ship name
Co-share code	Date of birth	Direction
Direction	First names (text)	Birth date
New Zealand port	Last names (text)	Intended length of stay/absence
Travel mode	Sex	Travel purpose
	Visa type code	Country of main destination
	Visa date	Country of last/next permanent residence
	Country of birth	Overseas state
	Route port code	Occupation
	Co-share code	New Zealand locality (city/ district)
Processing for statistical production		
Loading of movements and passenger data		
Validation		
Editing/correction		
Linking of imaged passenger card data		
Coding		
Linking of previous passenger classification history		
Derived variables:		
– Age		
– Length of stay/absence (days)		
– Passenger type (overseas visitor, New Zealand-resident traveller, permanent- and long-term migrant)		
Imputation		
Sampling and weighting		
Production of monthly statistics of passenger movements		
Series of linked travel histories for final definition of migrant status		
Passenger identifier		
Identifier linkable to other movement and passenger records (Identifier linkable to Statistics New Zealand Integrated Data Infrastructure)		
Direction of movement		
Date-time of movement		
Date of birth		
Derived long-term migrant status by '12/16 month' rule		

Figure 4.4 New Zealand usual resident count of overseas born population

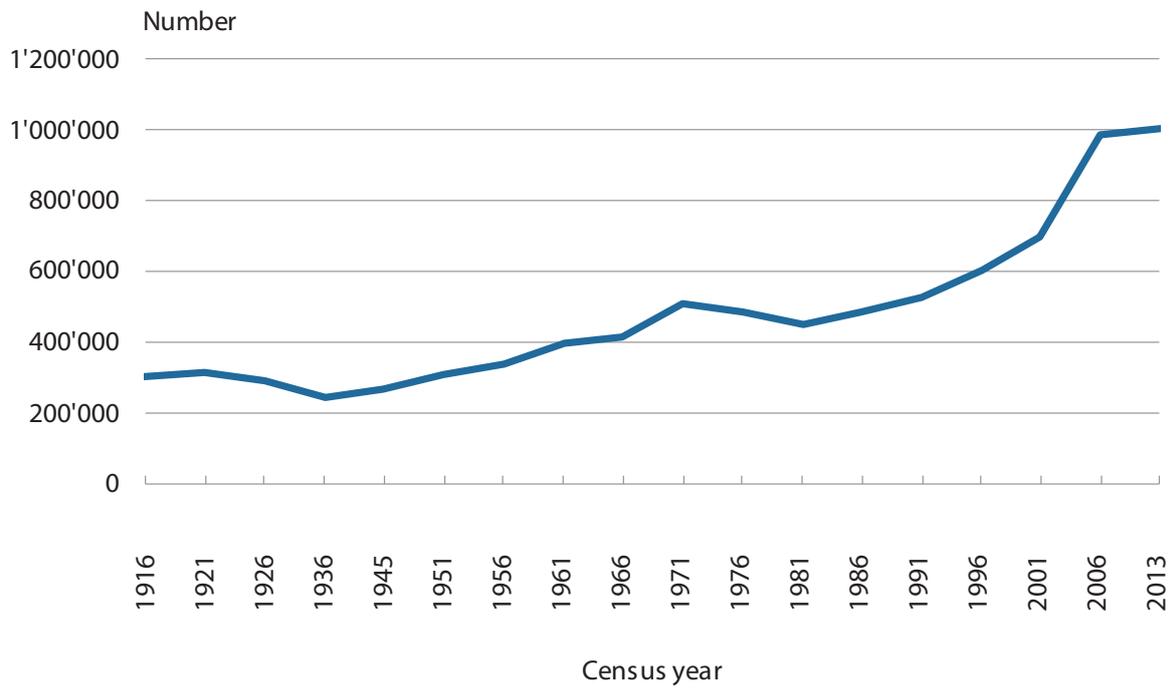
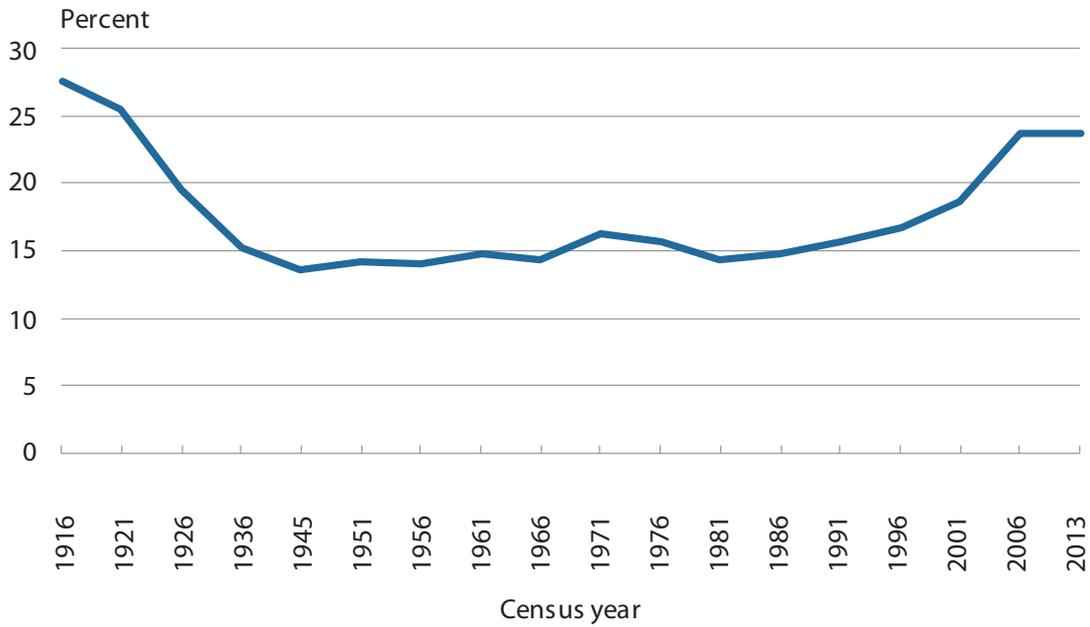


Figure 4.5 Percentage overseas born of New Zealand population



4.2.3.2 Purpose of data integration

4.2.3.2.1 Migration flows

246. The linking of imaged passenger card records to electronic border movements and passenger records is primarily undertaken to include additional passport information and carrier details available in the electronic capturing of border movements. The integration of border movements with passenger records also serves as a means of validation of the actual passengers with border movements at a given time. For some passengers there will be no filled-in cards, but electronic records will be available, or there may be filled-in cards but no electronic records. A manual linking process is included for optimizing the coverage of international passenger movements.

247. Linked travel and passenger classification sequences for the 16 months prior to the reference month are created for travellers who have indicated their intention to stay in New Zealand or to leave New Zealand long-term. The purpose is to confirm the New Zealand resident status at the time of travel using the information on the card about whether they have been living in New Zealand for 12 months or more. Given the integrated data sources and the classification history for the passenger, the evaluation serves as a data quality assurance check on initial counts of long-term arrivals and departures.

248. A passenger may change their intention about stay or absence after arriving in New Zealand or leaving New Zealand. An additional measure of migrant status takes account of when an international passenger is included in or excluded from the resident population after observing times of stay in New Zealand or absence from New Zealand. The measure uses linked travel histories over a follow-up period of 16 months and applies operational rules for determining a change in resident status. The '12/16-month rule' is independent of the individual's legal residence status and of the information stated on the passenger card.

249. By observing individuals' travel sequences over time, analysis of actual durations of stay or absence may produce results quite different from the information initially recorded on the passenger cards. A person may have transitioned from being counted as an overseas visitor arrival to be classified as a New Zealand resident traveller at a following overseas departure by going through stages of onshore temporary visa approvals to permanent residency. On-shore transitions leading to a change in residency status can result in over- or under-reporting of initial estimates of migrant arrivals and departures.

250. The use of travel histories also introduces added opportunities for extending international migration statistics. For example, there may be future interest in measures of short-term migration. Furthermore, passengers' classification histories represent a longitudinal data source enabling statistical measures of return and circular migration. Integration with the Statistics New Zealand Integrated Data Infrastructure (Statistics New Zealand, 2016) also facilitates other analysis of migration trajectories and outcomes.

4.2.3.3 Data integration methodology

251. Files of border movements and passenger data are automatically linkable through the record filing system provided by New Zealand Customs Service. The deterministic integration of these electronic files with imaged passenger cards includes an automatic step which links on passport number, name, and date of birth, carrier/ship name, and direction of travel. At least two fields from among passport number, name and date of birth must have full agreement for the record to be a match. A manual step is added to correct for non-linked records by looking at images of names and incorrect responses to the linking variables.

252. A small number of electronically recorded passenger movements have no associated filled-in passenger cards. Imputation for missing responses on the card is carried out using past travel history and passenger characteristics. Conversely, the passenger card may be the only source of border movement capture for some passengers. They mainly represent passengers arriving or departing by ship, and electronic records are created accordingly for these passengers.

253. The integration of these records with past 16-month passenger classification sequences is undertaken by deterministic linking on passport numbers, name fields and date of birth. Again, records are considered matched if there is full agreement on two of the three linking variables.

254. The deduplication of border movements to create the series of unique passenger IDs is a probabilistic process using the IBM QualityStage software. It applies the Fellegi-Sunter methodology (Statistics NZ, 2015) using key passenger identities; eg. name, date of birth, passport number. It assigns a unique passenger ID allowing for passengers changing passport nationality, using more than one passport, or for females changing their surname. For more information on the data integration methodology and the application of the '12/16-month' rule refer to the reference (Statistics NZ, 2017).

4.2.3.4 Use of data integration with other sources

255. Historic data sets of border movements and passenger classifications collected by the processes described above are stored within the Statistics New Zealand Integrated Data Infrastructure (IDI) (Statistics New Zealand, 2016). Travel and migration data in the IDI also include data sets containing information on the immigration approval decisions and criteria, as well as the temporary visa categories (e.g. student, work) granted on entry to New Zealand. The IDI research database contains integrated person-level data collected from a range of government agencies, Statistics New Zealand surveys including the 2013 Census, and non-government organizations.

256. Integration of travel and migration data with other sources such as the tax data allows longitudinal studies of migrants' contribution to the labour market. In particular, labour market outcomes may be compared across specific groups of migrants and for different geographic settlements.

4.2.3.5 References

Statistics New Zealand, International Travel and Migration, Data Collection Methodology, <http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/f705ca38-ea6e-453f-b1d9-a95dc0fc59#/nz.govt.stats/d83affa2-c0e1-4629-9aff-cc8452eb41c6#>

Statistics New Zealand, Defining migrants using travel histories and the '12/16-month' rule http://www.stats.govt.nz/browse_for_stats/population/Migration/define-migrants-travel-rule.aspx

Statistics New Zealand, Integrated Data Infrastructure http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx (2016)

Statistics New Zealand, Data integration manual <http://www.stats.govt.nz/methods/data-integration/data-integration-manual-2edn.aspx> (2015)

Statistics New Zealand, Coverage in the 2013 Census based on the New Zealand 2013 post-enumeration Survey http://www.stats.govt.nz/browse_for_stats/population/census_counts/report-on-2013-post-enumeration-survey.aspx (2014)

4.2.4 United Kingdom

257. The Office for National Statistics (ONS) produces statistics on international migration to and from the UK which contribute to our understanding of the make-up of society and the changing shape of the population. Our key migration statistics are:

- Long-Term International Migration estimates (flows);
- Short-Term International Migration estimates (flows);
- Population of the United Kingdom by Country of Birth and Nationality (stocks).

258. This case study focuses on the ONS Long-Term International Migration (LTIM) estimates.

4.2.4.1 Data sources for measurement of migration statistics

259. The ONS Long-term International Migration (LTIM) flows data are based on the International Passenger Survey (IPS) with several adjustments made based on other survey and administrative data sources.

260. The International Passenger Survey (IPS) is a multi-purpose survey that collects information from passengers as they enter or leave the UK. It is mainly used to provide data about international migration, travel expenditure and tourism.

261. The IPS is a sample survey which is conducted at all the main entry and exit points from the UK including airports, sea ports and at the Channel Tunnel.

262. To collect the data, a specific entry or exit point is sampled on a particular day. A strict method of counting is used which ensures each person has the same opportunity of inclusion in the survey. Passengers are systematically selected for interview at fixed intervals (for example, every 20th person) from a random start.

263. Between 700,000 and 800,000 people are interviewed on the IPS each year. Of those interviewed, approximately 4,000 people each year are identified as long-term migrants. These are non-UK residents who state that they intend to stay in the UK for at least 12 months (immigrants) or UK residents who state that they intend to reside outside of the UK for at least 12 months (emigrants). The remainder are identified as either short-term migrants, visitors to and from the UK, or UK residents returning after or leaving for a short stay overseas.

264. Further information on the IPS can be found in the International Passenger Survey Quality Information in Relation to Migration Flows.

4.2.4.2 Purpose of data integration

265. Although the IPS is the primary source of data on long-term migration, there is no single, all-inclusive system in place to measure all movements of migrants into and out of the UK. The IPS has some limitations with respect to measuring immigration and emigration, as it:

- Is a sample survey and so the estimates are subject to a degree of uncertainty;
- Captures very few asylum seekers who may be entering or leaving the UK, or non-asylum enforced removals from the UK;
- Does not take into account the changing intentions of passengers; these are passengers who intended to remain in or out of the UK for 12 months, but actually spent less than a year (migrant switchers) and those who believed they would be staying or leaving for less than a year but actually spent longer (visitor switchers);
- Does not capture those who are crossing the land border between the UK (Northern Ireland) and the Republic of Ireland.

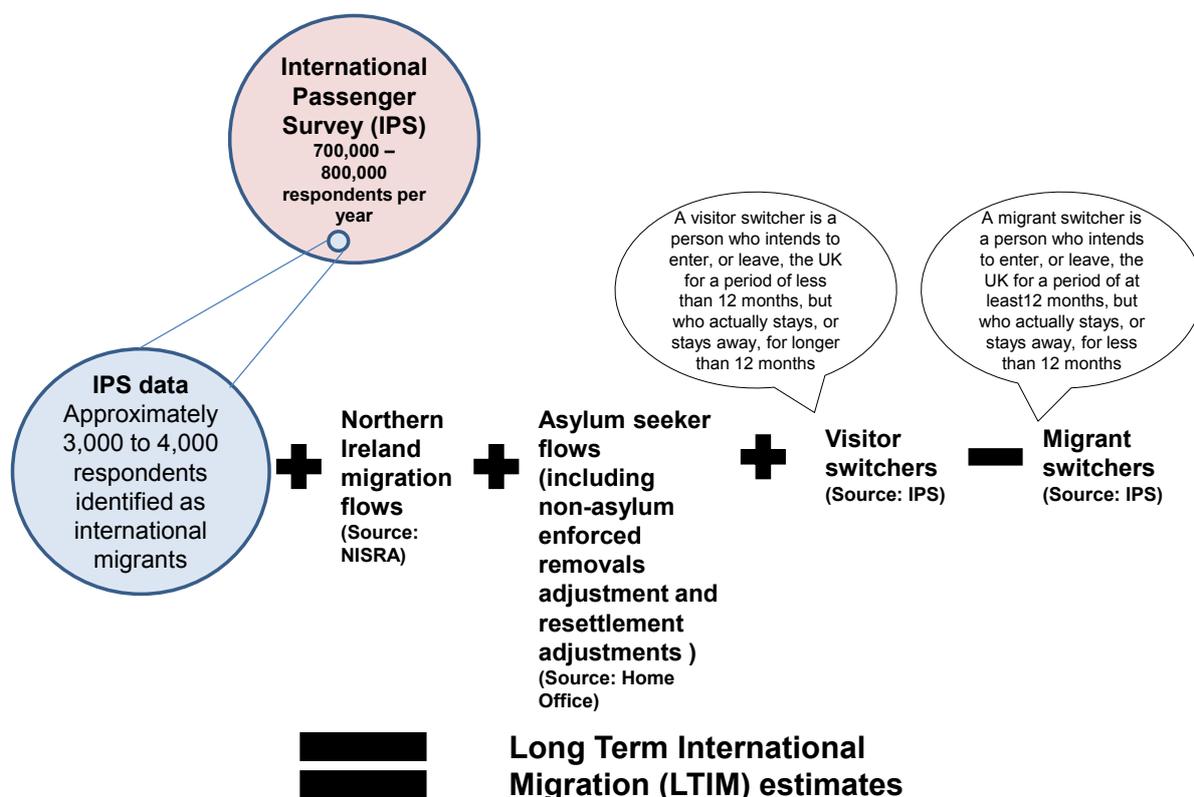
266. Therefore, it is necessary to use a combination of data from different sources that have different characteristics and attributes in order to produce estimates of LTIM.

4.2.4.3 Data integration methodology

267. The methodology outlined here was first applied in 2009 for the calculation of the 2008 estimates, and revisions were made to earlier years as appropriate.

268. Estimates of LTIM are about 90 per cent based on data from the IPS. A more comprehensive estimate of long-term international migration is produced by combining the IPS data with the information provided by additional data sources. Adjustments are also made for migrants who change their intentions (known as visitor and migrant switchers). This more inclusive estimate is referred to as Long-Term International Migration (LTIM) (see Figure 4.6).

Figure 4.6 Calculation of Long-Term International Migration



Source: Office for National Statistics

269. None of the data sources used, including the IPS, are specifically designed to capture information solely on long-term migration. Estimates of LTIM are produced using the following main data sources:

- International Passenger Survey (IPS);
- Labour Force Survey (LFS): provides a geographical distribution of migrants for the calibration of IPS;
- Home Office immigration administrative systems: provide data on asylum seekers and their dependants for 1991 onwards; from 2013 data on non-asylum enforced removals are also used; and from 2015, adjustments are made for people resettled in the UK under various resettlement schemes;
- Forecasted long-term international migration estimates based on previous General Practitioner (local doctor) registrations from the Northern Ireland Statistics and Research Agency (NISRA) for estimating long-term international migration to and from Northern Ireland and the rest of the world, from 2008 onwards (forecasted data is replaced with final data for LTIM final annual estimates). This captures land routes that are not surveyed by the IPS.

270. The published LTIM figures are broken down to show estimates by variables such as citizenship and age and sex. To produce estimates for each of these variables, data from the sources that contribute to LTIM also need to be broken down by the same variables.

271. Migrant data from the IPS are available broken down by each variable. Data on Northern Ireland flows and asylum seeker data are not, and need to be derived using a series of assumptions. In addition, the IPS data used to calculate the visitor switcher adjustments are based on a relatively small sample size each year, but still need to be broken down in the same way.

272. Since data received from the supplementary sources are not as detailed as the data collected by the IPS, the level of analysis that can be performed on the LTIM data is limited. While LTIM tables usually contain just one variable, cross-tabulations of IPS data are available separately.

273. LTIM estimates are shown with the confidence intervals for the IPS component of the estimate in order to give an indication of the accuracy of the estimate. The uncertainty associated with the IPS component of the estimate is used to calculate statistically significant changes in the LTIM estimates. However, when interpreting these confidence intervals and statistically significant changes, the reader should be aware that there is no method for quantifying the error associated with the non-survey components of LTIM, which are unlikely to be random.

274. Further information on the methodology, including the assumptions used to produce IPS and LTIM estimates²³, and changes to the methodology over time, is available on the ONS website. This also includes copies of the questionnaires used in the IPS.

4.2.4.4 Use of data integration with other sources

275. There are many sources of official statistics that measure the number and characteristics of international migrants into and out of the UK (flows) as well as the migrants who have settled in the UK (stocks). Taken together they provide a rich picture of migration in the UK. It is important to understand that these sources measure different things: some measure flows, some measure stocks, some measure workers, some students and some only measure the characteristics of those migrating from outside the EU. Each source is valuable in its own right in measuring particular aspects of international migration.

276. As noted, while offering the best data currently available, none of the data sources, including the IPS, is specifically designed to capture information solely on LTIM and therefore none of them taken alone can quantify all population inflows and outflows. Data integration is essential to ensure the best possible estimate of international migration is made, trusted and used confidently by the UK government and the wider community.

277. ONS is committed to providing the right information to inform public and policy debate. International migration is a topic of considerable debate and there is a high demand for trustworthy data and analysis. Regular methodological changes are made to ensure that estimates are as inclusive and robust as possible, particularly where new data sources become available. Details of changes can be found in the **Methodology to estimate LTIM**²⁴. In addition, in February 2017 ONS published a note on plans for the development of international migration statistics with a particular focus on making better use of current administrative data sources (see **International migration data and analysis: Improving the evidence**²⁵).

4.2.5 United States of America

4.2.5.1 Producing estimates of net international migration and its subcomponents

278. The US Census Bureau produces annual estimates of net international migration flows (NIM) to and from the United States, which are used as inputs for national and subnational population estimates. Integration of migration data is limited, although some foreign and administrative data are used to calculate a few subcomponents of NIM. At the same time, there has been a major effort at the Census Bureau to better leverage preexisting administrative data to reduce costs associated with Census data collection, although this has yet to be applied to international migration data. Some future work in this area with regard to migration data has been proposed, but is still at a nascent stage.

279. Data integration only occurs at the additive level, in that estimates of NIM subcomponents are derived from different data sources and then combined to produce a final estimate. The NIM estimate produced by the US Census Bureau is made up of several subcomponents which are calculated separately: 1) Foreign-born immigration, 2) foreign-born emigration, 3) net migration to/from Puerto Rico, 4) net native-born international migration, and 5) net military movement.

²³ "Methodology to estimate LTIM", <http://www.abs.gov.au/AUSSTATS/abs@.nsf/productsbyCatalogue/208F6F1B4DED24ECCA257EB50011C711?OpenDocument>

²⁴ <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/methodologies/internationalmigrationmethodology>

²⁵ <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/internationalmigrationdataandanalysisimprovingtheevidence/february2017>

The foreign born and Puerto Rico components utilize the American Community Survey (ACS) and its Puerto Rico equivalent. These subcomponents make up the bulk of the NIM estimates, but we also add a constant net native subcomponent that is derived from analysis of US-born or US citizens measured from foreign data sources (primarily censuses). The final subcomponent is net military movement to and from the US, which comes from administrative data supplied by the Department of Defense.

4.2.5.2 Movement of the Armed Forces Population to and from Overseas

280. The net military movement component uses the estimate of the net overseas movement of the Armed Forces population from data collected by the Department of Defense, Defense Manpower Data Center (DMDC). DMDC provides monthly tabulations of military personnel stationed or deployed outside the United States by age, sex, race, Hispanic origin, and individual branches of service within the Department of Defense. We assume that changes in the overseas military population, excluding deaths, indicate movement of personnel into and out of the United States. To derive estimates of net international movement of the armed forces at the state and county-level, we primarily use DMDC data by age, sex, race, Hispanic origin, and county. To improve the geographic distribution of military movement around certain domestic military installations, we use county grouping information derived from the most recent ACS five-year file.

281. This figure, at both the national and subnational level, is simply added to the other NIM subcomponents to produce a final NIM estimate. This is the only type of integration of international migration data we currently conduct.

4.2.5.3 Data integration at the micro-level

282. Integration of international migration data at the micro-level is currently not done at the Census Bureau, although we have experimented with looking at country of birth data from the Social Security Administration file (NUMIDENT) and merging them with Internal Revenue Service (IRS) tax data, to analyze domestic migration patterns of the foreign born. We have also discussed the potential of linking administrative data (from the Department of State) on refugees to census or ACS data in order to help determine refugee status through probabilistic methods. Research is also being conducted on using a combination of IRS (current residence), NUMIDENT (country of birth), and census data (Race/Hispanic Origin) to improve the subnational distribution of NIM estimates at the county level. Other administrative micro-level data exist at the Department of Homeland Security and their Office of Immigration Statistics (OIS). OIS has undertaken their own efforts to link various administrative data sets across Federal agencies, but at the US Census Bureau, we have only begun to think about possible ways to integrate migration data at these levels.

4.3 Summary

283. The case studies show a broad range of situations that differ widely with regard to the sources available for migration statistics, the limitations of those sources, and the data integration approaches adopted. The examples presented should not be considered as a fully representative sample of the data integration approaches developed in the various countries. Moreover, when evaluating the experiences of the various countries, it should be kept in mind that national contexts may differ significantly in a number of important aspects, and that a data integration approach adopted successfully in one country is not necessarily the best solution for other countries. Notwithstanding these limitations, the analysis of the case studies presented allows some key points to be made that could be useful for other countries.

284. In the **countries with a population register**, in several cases (including Austria, Italy, Latvia, and Spain) **this register is the main source of migration data**. In these countries data integration is carried out mainly to improve the data quality. Data from other registers can be used, among other purposes, to adjust for missing de-registrations of emigrants or to identify deceased persons. In several countries data integration is also used to provide additional information on variables not included in the primary sources.

285. Several countries adopt practices based on '**presence signals**' (in some countries called 'signs of life') of individuals in the various registers. In **Austria**, for instance, the analysis of presence signals in other registers is carried out to produce an estimate of those who are in the population register but no longer live in the country. Data from the Austrian Social Security administration are used to identify deceased persons, and data from various other administrative sources are used to adjust for missing de-registrations of emigrants.

286. In **Italy**, where only persons who are legally resident in the country can be included in the population register, the analysis of presence signals in the various sources allows improvement of the international migration estimates by identifying persons who are in the population register but have probably emigrated, and those who immigrated but are not in the population registers.

287. In **Latvia**, analysis of presence signals is carried out using various registers, including those on revenue, social insurance, education, health, and employment. Emigration estimates are also produced using mirror statistics from destination countries. Moreover, the quality of migration statistics is evaluated using data from household surveys, including LFS and EU-SILC.

288. In **Spain**, the national population register ('Padrón') is managed by the national statistical institute INE, and all persons living in the country – irrespective of their legal situation – have to be registered. In connection with the 2011 census, the registered population was linked with many other administrative registers in order to identify persons whose presence in the country was doubtful and those who were deceased. For doubtful records, data integration techniques based on administrative data and a sample survey carried out in connection with the census allowed estimation of the number of persons who were in the register but were not actually living in the country.

289. In some countries (including Hungary, Israel, the Netherlands and Switzerland), migration statistics are produced using **systems where the population register is one of several data sources**, not necessarily the most important one. In **Hungary**, a relatively complex system of registers is used, where the various sources are complementary to each other. Statistics on immigration and emigration are produced by integrating data from the various sources, in particular the data produced by the Office for Immigration and Asylum and the Central Population Register managed by the Ministry of the Interior. Data integration allows the production of high quality migration statistics and provides additional information to that available from the primary sources.

290. In **Israel**, data on migration flows and stocks derive mainly from border control registrations, and are provided by the Population and Immigration Authority (which also maintains the population register) to the Central Bureau of Statistics. The data are combined with other administrative files (including tax files, education files, etc.) and with data from the population census and other statistical surveys to provide estimates of greater quality and to add variables from other sources related to the same individuals or events.

291. In the **Netherlands**, migration statistics (like all social and demographic statistics) are based on the System of Social Statistical Databases, resulting from data integration among over fifty administrative registers. It is considered that virtually all the population living in the country is registered, with few exceptions. Concerning emigration, about one third of those leaving the country do not de-register. When the authorities find that someone is missing an investigation is conducted and, if not found, the person can be registered as emigrated to an unknown country.

292. In **Switzerland**, three main registers are relevant for migration data: 1) the local population registers maintained by the municipalities; 2) a federal register for legally resident foreigners holding a permit to stay; and 3) a special federal register for foreign diplomats, members of consulates, international organizations, their family members and similar categories. A modest form of data integration is carried out by matching data collected at the federal level and those collected at local level. Moreover, some variables available only in the register for legally resident foreigners are added to the local population register records pertaining to the same observations.

293. In the **countries without a population register**, the main sources of data on migration include passenger cards filled at borders, data on passports, visas and stay permits. As in the countries with population registers, data integration is carried out using various sources to improve the data quality and obtain more comprehensive information. In particular, data integration may help to improve the information on the actual place and date of settlement of immigrants, which often differ from those stated at the time of arrival.

294. In **Australia**, data on arrivals and departures are compiled by the Australian Bureau of Statistics based on information derived from various processing systems including passenger cards, passport and visa information. Data are linked using a personal identifier which allows ABS to also generate travellers' individual histories. Migrant data sets are also linked with census data and tax information.

295. In **Canada**, Statistics Canada receives administrative data on immigrants and non-permanent residents. Since the intended province and date of settlement do not always correspond with the actual place and date of settlement, data integration is carried out using other sources such as tax files. Record linkage is also carried out using data from other sources (such as the Canadian population census, the National Household Survey, the US Office of Migration Statistics) to provide estimates of greater quality and to obtain more comprehensive information on the socio-economic characteristics of migrants.

296. In **New Zealand**, imaged passenger cards are linked to electronic border movements and passenger records to include additional information and provide validation. Moreover, for travellers who have indicated their intention to stay in the country or leave it on a long-term basis, travel and passenger data are linked for the 16 months prior to the reference month, in order to confirm the status of New Zealand resident at the time of travel.

297. In the **United Kingdom** there are many sources of official statistics that provide information on different aspects of international migration flows and stocks. For flows, the main source is the **International Passenger Survey (IPS)**, a multi-purpose sample survey that collects information from passengers as they enter or leave the country. Since the IPS does not take into account the changing intentions of passengers, in order to produce long-term international migration estimates data integration is carried out using data from IPS, LFS, Home Office immigration administrative data, and estimates of the flows at the land border of Northern Ireland (not covered by the IPS).

298. In the **United States of America** the Census Bureau produces annual estimates of international migration flows by adding various subcomponents derived from different data sources. For some subcomponents a limited integration (at the aggregate level) is carried out using some foreign and administrative data, including the American Community Survey, foreign censuses, and military movements from the Department of Defense. Integration of international migration data at the micro-level is not done at the Census Bureau, although research is being conducted.

299. **To summarize**, the case studies show that virtually all countries covered use data integration to improve the quality of migration statistics, and to provide additional information to that available from the primary sources.

300. In countries with a population register, data integration – and the analysis of ‘presence signals’ – may allow the identification of:

- Persons who have emigrated but are still in the population register;
- Persons who have immigrated but are not in the population register; and
- Deceased persons.

301. In countries without a population register, data integration may help to improve the information on the actual place and date of settlement of immigrants, which often differ from those stated at the time of arrival.

302. The case studies showed many examples of data integration conducted at the micro or macro level. For example, in cases like Hungary or Canada different sources are micro-linked. This micro-integration is carried out by linking administrative sources on foreign people, both among themselves and with the population register, to avoid overlapping and to provide additional variables. As for macro-integration, an interesting example is found in the United Kingdom, where a border passenger survey is combined, by calibrating, with another source (Labour Force Survey) to allow territorial breakdown of incoming flows.

303. As expected, many different types of sources are used for data integration, including various administrative or statistical registers, other administrative data (including border control data), and statistical surveys (including censuses). Few countries use ‘mirror statistics’ from destination countries to improve their emigration estimates. This is done mainly at the aggregate level, although there are examples of international exchange of microdata, namely among the Nordic countries (not included in the case studies) and between countries that are neighbours.

5. Metadata

304. When migration estimates produced by different countries are compared, it is very important to have access to comprehensive metadata, in order to assess the comparability of the data in detail.

305. Ideally, three main types of metadata should be provided. The first is specific to each data source, describing its main features and detailing how the data sets are transformed before they are subject to integration; the second part looks at the way in which the data sets generated from the listed data sources are integrated; the third part would be based on some measures of quality, demonstrating the increased informative content resulting from integration. This information should be provided for all relevant migration statistics, i.e. for migrant stocks, migration flows, and any sub-groups of interest.

306. The first information that should be provided is the list of data sources used to produce the integrated data. In such a list, the following metadata could accompany each data source:

- Name;
- Type (administrative register, sample survey, census, big data, etc.);
- Owner (including statistical authority of another country);
- Rules of access / data provision (if the owner is other than NSO)
- Population of reference;
- Date/period of reference of the data;
- Frequency of updating;
- Timeliness (i.e., the length of time between the date of availability of the data and their date/period of reference);
- Level of detail of the original data (micro, macro);
- Level of detail of the data set used for integration (micro, macro);
- Description of any method applied to transform the original data set into an input suitable for data integration;
- Dimensions of the data set (n x p);
- The variables contained in the data source (or broad description if there are many);
- The variables contained in the data sources that are retained for integration (possibly subset of the previous item);
- The variables used to integrate this data source with the other(s) (possibly subset of the previous item);
- Availability of the PIN (this is in principle included in the previous item, but given its relevance it could be useful to dedicate a specific space to it) and its main features where relevant.

307. Assuming that the integration process is implemented sequentially, for each step, i.e., for each pair of data sets being integrated, the following information could be provided:

- Step number;
- Data sets involved (using names or number from the list of individual data sources);
- Frequency of the integration procedure;
- Variables used for integration (if applicable);
- Description of the methodology for integrating the data sets;

- Share of data overlapping;
- Main issues or difficulties in the integration of these data sets, such as reporting of non-linkage, methods and results from estimation of false-positive link rates;
- Dimensions ($n \times p$) of the resulting integrated data set.

308. In some cases, and particularly when migration statistics are derived from population registers, the number of data sources may be relatively high, and thus the reporting of the metadata can become burdensome. In such cases, it may be suitable to prepare a single general report on the functioning of the system. However, it should also be taken into account that it may be important for the users to have enough information to form a clear understanding on how the final data are produced through a possibly complex migration processing system. Complexity should not preclude transparency.

309. The third part of the metadata, which requires much further work, can be developed around the following initial set of quantitative measures of quality, aiming to highlight the positive effects of data integration:

- Difference in the number of records or the size of the population covered, between the final integrated data set and the smallest single data set: $n^* - \min(n)$;
- Difference in the number of variables, between the final integrated data set and the smallest single data set: $p^* - \min(p)$;
- Difference in timeliness, between the final integrated data set and the least timely single data set: $t^* - \max(t)$;
- Difference in the number of records or the size of the population covered, between the final integrated data set and the largest single data set: $n^* - \max(n)$;
- Difference in the number of variables, between the final integrated data set and the largest single data set: $p^* - \max(p)$;
- Difference in timeliness, between the final integrated data set and the most timely single data set: $t^* - \min(t)$;
- Gain in the quality of estimates, such as percentage decrease of the variance of the estimates, or any quantitative measure of the improved accuracy.

310. The metadata can be enriched by a qualitative assessment of the impact of integration, such as feedback given for the improvement of the original data sources and reduction of response burden, possibly supported by specific references or examples, and by an overall conclusion about the integration process.

6. Conclusions

311. Migration statistics are probably the most complex element in the field of social statistics, not only from an operational point of view, but also from a conceptual point of view. While it is true that there are international definitions, it is not always easy to measure the strict concept of a migrant. Moreover, there are increasingly different forms of migration, and there is a variety of sources that provide partial measurement of migration.

312. The challenge for statistical offices is to combine data from the different sources into reliable and integrated information on migration. The 13 case studies showed the different ways and data sources that countries use to improve their migration statistics.

313. The Task Force conducted a survey to assess the potential of many different sources for producing migration statistics. According to the replies, most countries (31 of the 56 responses) use **more than one source for generating migration statistics**. Approaches differ because the circumstances and data sources vary greatly across countries. The overall tendency to involve more administrative sources in measuring migration applies both to countries that base their population statistics on classical methods (surveys, exhaustive censuses) and to countries that base their statistics on population registers.

314. The choice of sources for migration data often depends on **geography**. Countries that are relatively isolated or with fewer points of entry and exit (e.g. New Zealand, Malta) can produce migration statistics based on very different sources than others with many access points or without checkpoints, such as countries in the Schengen Area in Europe.

315. Another important limitation when studying migration statistics is that there are **different definitions** of the migrant population in different administrative sources. In many cases, the standard time frame of 12 months of actual or planned stay is not the one used for the preparation of migration statistics, but other concepts (e.g. other time periods) are adopted. In addition, different administrative sources within the same country can offer information based on different concepts.

316. In the European Union countries, migration statistics are intended to be fully **consistent with changes in migrant stock**. In many other cases, migration statistics do not pursue this full consistency between flows and stocks. Moreover, many countries produce separate data on long-term migrants, asylum seekers, non-permanent foreign population, short-term migrants, etc.

317. There are some relevant examples of integration of data **at both macro and micro levels**. As regards micro-integration, that is, the combining of sources at the individual record level, there are several examples that may illustrate the trend for the future. The combining of records may generate overlaps and these are addressed through adjustment and calibration processes. This integration takes place to improve the measurement of stocks and also to produce more complete statistics on flows.

318. The main weakness of the population registers is their **difficulty in detecting emigration**. If a person does not inform the authorities of their exit from the country, the administrative register has to be updated by administrative procedures that are not always easy or immediate, so the population registers may contain persons no longer residing in the country.

319. By combining sources, '**presence signals**' can be detected. For example, a population register can be linked with tax or social security files to be more certain about the presence of people in the country. This is of particular importance in the combination of different administrative records in cases such as Latvia or Austria, where logistic regressions are used for estimating resident population from evidence contained in different administrative registers. The analysis of the 'presence signals' may also help identify persons who immigrated but are not in the population register (as in Italy), and deceased persons (as in Austria and Spain).

320. As for statistics on migration flows, there are some examples of integration in the strict sense, i.e., the combination of sources to obtain a single database. For example, in Hungary and Canada different sources are micro-linked. This micro-integration is carried out by linking information on foreigners across different administrative sources and with the population register, to avoid overlapping and to provide additional variables.

321. An initiative that is beginning to be demanded from users and that would be very promising is **the exchange of individual-level data** between countries for statistical purposes. A fundamental aspect of migration flows is that they affect one country of origin and another of destination. So it is logical to think that the sharing of data will improve these statistics. The exchange of individual-level data for statistical purposes is likely to develop further, although it is necessary to overcome legal, technical and other barriers. Nordic countries provide a good example of such exchange.

7. Recommendations

322. Based on the work carried out by the Task Force and the content of this document, a number of recommendations to national statistical offices are presented in this section. These recommendations are considered relevant to all countries, regardless of their current situation with the availability and use of administrative data for migration statistics.

7.1 Improve access to administrative data for national statistical offices

323. In countries where administrative records are not available or only partly available to the statistical offices to produce official statistics, a priority is to take measures to facilitate the access to administrative records and to promote their use by the statistical office. This may require support at the political level and, in some cases, changes in the legislation.

7.2 Use administrative data for migration statistics

324. Many administrative sources contain information that can be used in migration statistics. These include registers and databases on population, visas, passports, border control, residence and work permits, and asylum seekers and refugees. It is recommended that the potential of all these administrative sources for migration statistics be examined, and that they be used for the production of migration statistics to the extent they are fit for this purpose. This requires cooperation with relevant ministries, agencies and administrations responsible for the data.

325. It is advisable to produce several statistical outputs that provide information on migration stocks, as well as inflows and outflows, covering different types of migration flows. In some cases different statistics may not provide fully consistent data. However, their publication can help to give a global picture and to increase awareness that a single product does not provide all the necessary information on international migration. The simple publication of data from various sources is a first mechanism for integrating information on migrants.

7.3 Combine data from different sources using ‘presence signals’

326. In those countries where administrative data are available for statistics, the approach of combining sources using ‘presence signals’ methods is very promising. The defining feature of such approaches is the use of several sources to estimate the probability of an individual’s residence in the country based on their presence in these sources. Combining data from different sources in such a way permits the improvement of estimates of current population and stock of foreigners.

7.4 Pay attention to the quality of integrated data

327. As countries advance and standardize registers of social and population data, data integration is more likely to take place at the unit record level across multiple sources. Greater attention to quality assessment of the integrated data will necessarily become an integral part of the production of international migration statistics.

7.5 Be transparent about data integration methods used and develop standards

328. It is increasingly important for statistical offices to provide a transparent presentation of the rules and methods used for the data integration processes that produce the statistics on international migration flows and other types of international migration statistics. This opens up opportunities to develop methodological standards for data integration practices across the national statistical offices in the future. Individual countries will always have their own collection and data integration practices but there are likely to be more open technical forums for discussion of the rules and methods used in the combining of sources for the production of migration statistics.

7.6 Promote international comparison and exchange of migration data

329. Another source for improving migration statistics is the international comparison and exchange of data on migration, foreign-born population and foreigners. Administrative sources typically underestimate emigration due to the low motivation of emigrants to register the event of moving out of the country. However, the immigration data of countries where those emigrants go could cover these events. This can be checked using the so-called 'mirror statistics' that show the migration flow from country A to country B through the lenses of emigration statistics of country A and immigration statistics of country B. Exchanging data with other countries can thus be very useful for understanding migration processes and for improving their systems for measuring migration.

330. International data sources are very helpful for this. In the European case, Eurostat publishes quite detailed data of inflows and outflows that allow this comparison between countries of the European Union. In its Clearing House (database) of Migration Statistics, UNECE publishes such data from countries of Eastern Europe, Caucasus and Central Asia.

331. A breakthrough could be achieved if the databases were shared between countries to link at the individual level. There are already cases where this occurs, such as in the Nordic countries. At present it is unlikely that exchanges of individual-level data will become the norm, but there might be some initiatives between countries due to the intensity of their exchanges or others circumstances. The success of these exchanges could encourage others to follow the same path.

8. Further work

332. At the international level, further work could address the potential of using big data for migration statistics and the ways in which the analysis of longitudinal data could be incorporated into annual migration statistics to complement the available cross-sectional measurements.

8.1 The potential of big data

333. New and non-conventional data sources, such as data gathered from the use of mobile telephones, credit cards and social networks — generally known as big data — could potentially be useful for producing migration statistics. Currently, statistical offices have great difficulty in accessing these data, which are in most cases privately owned. In the coming years, steps should be taken to gain access to these sources and to find ways to use them for the purpose of producing official statistics. This information alone would probably be insufficient and biased, but when integrated with conventional sources it could improve migration estimates. It will be important to share the emerging practices internationally, to support countries' first steps towards harnessing the potential of such data for producing migration statistics.

8.2 Longitudinal data

334. The future of data integration probably does not lie only in the improvement of quality in the measurement of migratory flows. The integration of data may lead to the availability of new information on the migrant population. The possibility of connecting population databases (or surveys) with files for employment, social security, taxes, health, and education systems, among many others, will provide more and better information in the future on living conditions or the social integration of the migrant population.

335. A clear example is the potential offered by longitudinal data. At present, official statistics focus almost exclusively on providing cross-sectional information on migrants. For example, barely any information is gathered by statistical offices about the duration of the migratory episodes (does a migrant live in the host country for a long time?) or about time intervals —which can be of great relevance when studying migratory flows. For instance, information about the time elapsed between arrival in the country of destination and obtaining a job, or between losing a job and emigrating, could be very informative. Longitudinal data are needed to address the question of whether migrants' socio-economic outcomes improve or worsen over time. They are also needed to understand family migration and emerging patterns such as circular migration.

336. Sharing of good practices in creating data sets with a longitudinal perspective and in using them for regular migration statistics would support NSOs in their work. The next step would be developing recommendations on how to harmonize such statistics internationally.

References

- ESSnet (2008): Project on Integration of Survey and Administrative Data. Material available at: https://ec.europa.eu/eurostat/cros/content/isad-0_en
- ESSnet (2011): Project on Data Integration. Material available at: https://ec.europa.eu/eurostat/cros/content/data-integration_en
- ESSnet (2014): Project on Handbook on Methodology of Modern Business Statistics ('Memobust handbook'), Module on Macro Integration. Material available at: https://ec.europa.eu/eurostat/cros/content/macro-integration_en
- Eurostat (2009): "Insights on Data Integration Methodologies". Proceedings of the ESSnet-ISAD workshop, Vienna, 29-30 May 2008. Eurostat Methodologies and Working Papers. Available at: <http://ec.europa.eu/eurostat/en/web/products-statistical-working-papers/-/KS-RA-09-005>
- Eurostat (2013): "Statistical matching: a model based approach for data integration". Eurostat Methodologies and Working Papers. Available at: <http://ec.europa.eu/eurostat/en/web/products-statistical-working-papers/-/KS-RA-13-020>
- Sheps M. (2016): "Challenges in estimating irregular migration in Israel since 1995", Working paper 14 for the Joint UNECE/Eurostat Work Session on Migration Statistics, Geneva, Switzerland, 18-20 May 2016. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.10/2016/mtg2_WS/14_Israel_Sheps_ENG.pdf
- Sheps M. and Hleilel A. (2006): "The challenge in creating a stock of emigrants from Israel". Working paper 10 for the Joint UNECE/Eurostat Work Session on Migration Statistics, Edinburgh, Scotland, 20-22 November 2006. Available at <https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.10/2006/wp.10.e.pdf>
- SDMX (2009): "SDMX Content-Oriented Guidelines – Annex 4: Metadata Common Vocabulary". Available at <http://www.sdmx.org>.
- UNECE High-Level Group for the Modernisation of Official Statistics (2017): "In-depth review of data integration". Document ECE/CES/2017/8 for the Conference of European Statisticians meeting of 19-21 June 2017

Guidance on Data Integration for Measuring Migration

Migration shapes societies. Its economic, social and demographic impacts are large and increasing. Policymakers, researchers and other stakeholders need data on migrants – how many there are, their rates of entry and exit, their characteristics, and their integration into societies. These data need be comprehensive, accurate and frequently updated. There is no single source that can provide such data, but by combining several sources together it might be possible to produce the information that users need. Some countries have developed methods for combining administrative, statistical and other data sources for the production of migration statistics.

This publication, developed by a task force of experts from national statistical offices and endorsed by the Conference of European Statisticians in June 2018, provides an overview of the ways that data integration is used to produce migration statistics, based on a survey of migration data providers in over 50 countries. Thirteen case studies provide more detail on data integration in various national contexts. The publication proposes principles of best practice for integrating data to measure migration.

The publication is designed to guide national statistical offices and other producers of migration statistics, while also offering users an insight into the production of the migration data they use.

Information Service
United Nations Economic Commission for Europe

Palais des Nations
CH - 1211 Geneva 10, Switzerland
Telephone: +41(0)22 917 12 34
E-mail: unece_info@un.org
Website: <http://www.unece.org>

ISBN 978-92-1-117185-3



9 789211 171853