



Economic and Social Council

Distr.: General
4 April 2022

Original: English

Economic Commission for Europe

Conference of European Statisticians

Seventieth plenary session

Geneva, 20–22 June 2022

Item 6 (d) of the provisional agenda

Reports, guidelines and recommendations prepared under the umbrella of the Conference:

Use of new data sources for measuring migration and cross-border mobility

Use of new data sources for measuring international migration

Prepared by the Task Force on the use of new data sources for measuring international migration and cross-border mobility

Summary

This report presents the results of the work of the Task Force on the use of new data sources for measuring international migration and cross-border mobility, composed of representatives from Italy (Chair), Canada, Georgia, Mexico, New Zealand, Turkey, United Kingdom, United States, Eurostat, the Organisation for Economic Co-operation and Development (OECD), the United Nations Economic Commission for Europe (UNECE), and the United Nations Statistics Division (UNSD).

The report presents national experiences with big data and new data sources collected through two surveys among countries participating in the work of the Conference of European Statisticians (CES). The report concludes that national statistical offices have limited experiences in the use of new data sources for migration statistics. However, the potential of new data sources to address emerging migration issues and urgent data needs is promising. Sharing and updating of relevant information and tools will facilitate the use of big data and new data sources for measuring migration and cross-border mobility.

The current short version of the document is prepared for translation purposes and – compared to the full version – does not contain information on the distribution of articles included in the DIMiS database by geography and data sources (at the end of chapter 5).

The full text of the document was sent to all members of the Conference for electronic consultation in March–April 2022 and is available at the webpage of the 2022 plenary session at <https://unece.org/statistics/events/CES2022>. Results of the electronic consultation will be summarized in document ECE/CES/2022/6/Add.1.

Subject to a positive outcome of the consultation, the CES plenary session will be invited to endorse the document.



I. Introduction

A. Background

1. Migration and other forms of cross-border mobility are issues of high policy importance. Demands for statistics in these areas have further increased in light of the [2030 Agenda for Sustainable Development](#) and the [Global Compact for Safe, Orderly and Regular Migration](#) (2018). The statistical community continues to be challenged to capture international migration and cross-border mobility in a way that would meet the growing needs of users.
2. Measurement of migration and cross-border mobility relies on a variety of sources, such as population and housing censuses, household surveys and administrative records, with each of them having their own strengths and limitations. Integration of data from different sources is a way to enhance the richness of data and reduce coverage or accuracy problems. However, even this will often not capture all dimensions of migration and cross-border mobility.
3. New non-conventional data sources, such as data gathered from the use of mobile telephones, credit cards and social networks — generally known as big and social media data — could be useful for producing migration statistics when used in combination with conventional sources. The [UNECE Guidance on data integration for measuring migration](#) proposes further work on utilising the potential of big data, “to share the emerging practices internationally, to support countries’ first steps towards harnessing the potential of such data for producing migration statistics”.
4. Notwithstanding the challenges of accessibility, accuracy and privacy of these new sources, examples have emerged in recent years that highlight their potential. The [2018 UNECE-Eurostat Work Session on Migration Statistics](#) illustrated the use of Facebook data for obtaining age profiles of ‘expats’ by origin and the use of geo-tagged tweets for estimating mobility. The [2019 UNECE-Eurostat Work Session on Migration Statistics](#) featured an example from official statistics, from the United States Census Bureau, on the use of air passenger data for improving migration estimates. As more examples emerge, they need to be collected and analysed, to support national statistical offices (NSOs) in embarking on the use of new data sources, building on the results of the [UNECE Big Data Projects for Official Statistics](#), and other initiatives that look at the relationship of official data providers with big data owners.
5. In October 2019, the Conference of European Statisticians (CES) Bureau reviewed in-depth the statistics on international migration and cross-border mobility, based on a paper by Mexico¹ and a note by UNECE. The Bureau brought up the importance of gathering examples where NSOs are using the new data sources for producing official statistics in this area and pointed at the 2019 UNECE-Eurostat Work Session on Migration Statistics as a pertinent forum to discuss this further. It requested the Secretariat and the Steering Group on migration statistics to present to the next Bureau meeting a proposal for further work on the use of new data sources for measuring migration and cross-border mobility.
6. Participants of the 2019 UNECE-Eurostat Work Session on Migration Statistics recognised the need to work towards using new types of data sources, such as mobile devices, social media networks, satellite images and Internet platforms, and to review existing examples of use of new data sources for the benefit of producing official migration statistics. To meet this need and the Bureau’s request, the CES Bureau in January 2019 established the Task force on the use of new data sources for measuring international migration and cross-border mobility to review existing experience and plans in NSOs for using new data sources for measuring migration and cross-border mobility, identify relevant examples from outside official statistics, compile the examples into a reference tool and develop a mechanism for

¹ http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/bur/2019/October/02_In-depth_review_on_international_migration_Rev.1.pdf

updating it with new examples, and analyse the collected material to guide NSOs in the use of the new sources.

B. Methodology and content of the report

7. The Task Force started its work identifying relevant scientific literature and the information available among its members. To collect information on relevant experiences (in NSOs or outside) on the use of new data sources for measuring migration and cross-border mobility, the Task Force carried out between November 2020 and February 2021 an online survey among NSOs in the UNECE region.

8. Out of the 34 UNECE countries that participated in the survey, only 5 (15 percent) reported that they were aware of a process or a project (whether or not involving the NSO) that uses new data sources, such as big and social media data, for measuring migration and cross-border mobility.

9. The Task Force analysed the results of the survey and decided to carry out a follow-up survey to collect from NSOs additional information on:

(a) Whether NSOs make use of new data sources for producing statistics in other statistical fields;

(b) The main reasons why NSOs do not make use of new data sources for migration statistics;

(c) Whether NSOs are working on any significant innovations on measuring international migration, including new approaches using current sources, or exploring new data sources.

10. The follow-up survey was carried out in March 2021, and 27 countries responded.

11. The results of the two surveys are presented respectively in the sections II. National experiences with big data and new data sources, III. Reasons for not using new data sources, and IV. Significant innovations on measuring international migration.

12. In addition to the information provided by NSOs in the surveys, the Task Force also collected a large amount of information on different types of research, discussion papers or scientific journals related directly or indirectly to the use of new data sources for measuring international migration and cross-border mobility. The Task Force developed a concept for an online reference tool to present this material, with the possibility to update the tool with new examples that would become available in future, as it was requested by the CES Bureau. The synopsis developed by the Task Force for the reference tool is presented in “ V. A literature reference tool”.

II. National experiences with big data and new data sources

13. Based on the two surveys conducted by the Task Force, it is evident that NSOs rarely use big data currently to measure international migration. However, some agencies are using new data sources, and others are exploring the use of new data sources.

A. Summary of experiences: first survey

14. The one reported project using big data is in Georgia. The project aims to measure human mobility, as part of the deliverables of the United Nations Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD)². The task is to estimate population mobility patterns such as internal migrants, international migrants and tourists. The NSO in Georgia (Geostat), the Georgian mobile phone regulator (GNCC) and other members of the task team (International Telecommunication Union (ITU), UNSD, Eurostat,

² <https://unstats.un.org/bigdata/blog/2019/mpd-task-team.cshtml>

Positium, and others) are working together to develop and test methods and methodology to estimate migration and tourism statistics in Georgia with use of mobile phone data. However, the project is yet to produce any official statistics and is currently on hold for technical reasons.

15. One example of using new data sources, rather than big data, is reported by the United States Census Bureau. This project is based on using passenger flight data to adjust net international migration between Puerto Rico and the United States, to include the impact of Hurricane Maria in 2017. The United States Census Bureau has made a concerted effort in recent years to integrate new data sources to improve its method to estimate net migration between the United States and Puerto Rico. These efforts have been motivated by a need to improve the timeliness of migration estimates which have typically relied on large household survey data collected by the Census Bureau. A survey-based methodology works well when migration patterns are consistent over time but tends to perform less well when migration patterns quickly change, such as due to natural disasters (e.g., hurricanes, global pandemics). To provide more up-to-date and accurate estimates of migration, the Census Bureau has worked to integrate administrative data produced by other agencies with data collected from its surveys, while moving towards a solely administrative data-based method to measure net migration between Puerto Rico and the rest of the world.

16. The impetus for these initial efforts to combine survey and administrative data was a result of the impact Hurricane Maria had in 2017, which caused mass out-migration from Puerto Rico. These macro-data integration methods combined flight data published by the Bureau of Transportation Statistics (BTS) and survey-based estimates from the American Community Survey (ACS) and Puerto Rico Community Survey (PRCS) to better measure the effect of this natural disaster for 2018 estimates. Macro-data integration was again used to account for return migration to Puerto Rico after Hurricane Maria (2019 estimates), as well as adjust for new migration patterns resulting from the COVID-19 pandemic (2020 estimates). Starting in 2021, the Census Bureau is adopting a method which produces total net migration for Puerto Rico using flight data, but still relies on the ACS and PRCS to develop inflows and outflows by demographic characteristics. This new method is advantageous since it will improve the timeliness of data availability, reduce the need for adjustments, align better with the reference period (year ended 30 June), and will expand the migration data to include international moves to and from Puerto Rico (for details see C. Example: the United States experience).

17. Other projects were indicated as examples of the use of big data, but related to internal migration (e.g. Netherlands) or the results are not yet available (e.g. Latvia).

B. Summary of experiences: second survey

18. In the follow-up survey, four countries reported that they are exploring the possibility of using new data sources for measuring international migration.

19. Just one project is based on what might properly be called ‘big data’. The United Kingdom Office of National Statistics is currently exploring aggregated and anonymised mobile phone data (or crowd movement data) from O2 Motion. These data look at mobile device connections to O2 masts to gain a picture of crowd movement trends around the UK. The data includes geographic location (where mobile phones are connected to cell phone towers) at national and local authority granularities and has the potential to help understand patterns of international migration. The plan is to assess what the data says about mobility in and out of the United Kingdom, including aspects such as duration of stay, as part of the work looking for timelier indicators of migration. This project is at a very early stage and there is a wider program of work for international mobility within the United Kingdom's data science campus that is exploring more than this³. Air passenger information is also used as one of several sources for modelling migration.

³ Data insights from O2 Motion never allow identification or mapping of individuals and operate within strict privacy guidelines. [Using Mobile Phone Data for Enhancing International Passenger Survey Traveller Statistics](#)

20. Other countries are exploring new data sources but not necessarily big data.
21. Statistics New Zealand reported that administrative data (passport information from border-crossings) is used for official measures of international migration but also other data sources – such as mobile phone data – are being explored for studying geographic population distributions within New Zealand. This includes international visitors, and residents (including recent migrants).
22. The Statistical Office of the Slovak Republic does not use big data for measuring migration statistics but are focusing on the innovative use of administrative data. The systematic use of administrative data sources for official migration statistics is described as the main challenge.
23. Statistics Lithuania reported that the application of mathematical methods to new data sources (and big data) could help overcome challenges related to international migration statistics. A concurrent development in Lithuania is the State Data Governance Information System, providing for a common data management platform across agencies and enabling data accessibility for state needs, efficient exchange between institutions, and data sharing with the business and science community. This system will increase the potential for using government information systems, registers, and other sources for international migration statistics.
24. The Italian National Institute of Statistics is also working on the integration of new data sources for migration statistics. In Italy the international migration statistics are based on the migrants' notification to the population register of their arrival or departure. A specific research project aims to make use of the registrations and de-registrations ex-officio to estimate the missing part of these notifications, especially for emigration statistics, as it is well known that emigrants might not communicate their departure. In some cases the de-registration ex-officio is followed after a few months by a registration for reappearance of the same individual. The Italian longitudinal system (ANVIS), based on a micro-demographic accounting, allows identification of these offsetting administrative movements and removes them from the demographic balance. In contrast, it is likely that those who don't reappear have left the country. In some cases (i.e., the time between the deregistration ex-officio and the registration for reappearance is more than one year) there is uncertainty whether the person has left the country and then returned or has always resided in the country during all of the period. To this aim the presence of these individuals is checked against another Integrated Archive of Usual Resident Population that includes and integrates many administrative sources (Labour and Education registers, Tax Returns register, Earnings, Retired, and Non-Pension Benefits registers, Permits to Stay archive). This archive, already used for the assessment of coverage of the census resident population, could be used also to indicate presence in Italy by using individual's "signs of life". That is, checking the hypothesis that the person has left the country by looking at his or her signs of life in the period.

C. Example: the United States experience

25. This section focuses on the United States Census Bureau work to incorporate new data sources to measure migration, particularly for estimates of Puerto Rico migration. These new data sources and methodologies have helped to overcome limitations with previous methods dependent on more traditional data sources. By integrating new and old data sources, timelier and more accurate estimates of migration have been produced. The Census Bureau will continue to pursue these avenues of research to improve estimates.

1. Traditional survey sources

26. The United States Census Bureau produces annual estimates of the population as of 1 July each year. The American Community Survey and Puerto Rican Community Survey (ACS/PRCS) are annual continuous household surveys of the United States population that ask detailed information previously collected on the decennial census long form. The ACS currently surveys about 3.5 million households per year, while the PRCS has a sample of 36,000 Puerto Rican addresses.

27. Using the previous methodology, estimates of migration flows from Puerto Rico to the United States were based on responses to the ACS residence one-year ago (ROYA) question, which asks where respondents lived one-year before the survey. Conversely, migration flows from the United States to Puerto Rico were derived from the PRCS ROYA question. The net migration estimate is derived from subtracting the in- and out-migration flows.

28. Data are collected on a continuous basis throughout the calendar year, though movement could have occurred at any time over a 2-year period, depending on when the respondent was included in the sample and when they actually moved. Puerto Rico net migration not only impacts the population estimate for Puerto Rico, but also for the mainland United States, as it is included as part of the mainland's net international migration component.

2. Air passenger traffic data as an alternative source

29. In the United States, flight data are compiled from monthly reports filed by over 200 commercial United States and foreign air carriers with the BTS, including both domestic and international flights. Monthly domestic flight data are released with about a 3-month time lag. Release of annual data on international flights lags about 6 months from the end of the year, with complete international data available in the middle of June of the following year. Data are reported for all flights (thus no sampling is involved) following federal reporting guidelines which went into effect in October 2002.

30. For Puerto Rico, air passenger traffic (APT) domestic data give monthly information on the number of passengers flying on planes between Puerto Rico and the mainland United States. It should be noted that APT data include information on all travellers without differentiation of passenger type, and thus includes tourists and visitors who make up most passengers. Non-migrants are counted on both their in-bound and out-bound flights, while migrants are only counted in one direction – unless leaving temporarily, in which case they would be counted again upon their return.

31. A limitation of this method is that it can only give a number for “net” migration, with no information on total inflows or outflows, as migrants cannot be distinguished from the total number of passengers entering or leaving Puerto Rico. Another limitation is that no demographic characteristics of migrants are included in the data. Additionally, this method is only applicable to a country or territory without any land borders, such as an island like Puerto Rico, as flights are the main method of arriving or leaving. Ship passenger movements to and from Puerto Rico are assumed to be minimal.

32. Monthly tallies of net airline passenger flow movement reflect seasonal variations related to tourism, with greater movement into and out of Puerto Rico in the summer and winter vacation months. Depending on the measurement period (e.g., a calendar year), this could lead to year-to-year fluctuations related to annual tourism trends. For example, a high number of tourists could be counted in December, while the return of these same tourists might happen in January of the following year. Over time, these fluctuations are thought to balance out.

3. Hurricane Maria

33. The Commonwealth of Puerto Rico is an unincorporated territory of the United States with a population of over 3 million people. Puerto Rico's population has been declining since 2004, primarily due to out-migration to the United States, coinciding with economic decline on the island. Puerto Ricans are United States citizens and have the right to free movement between Puerto Rico and the United States, resulting in over 5 million people of Puerto Rican origin living in the United States.

34. In September 2017, Category 5 Hurricane Maria made landfall on Puerto Rico, resulting in extensive damage, loss of human life, and out-migration to the mainland United States. Natural disasters can impact population, namely through the movement of persons from affected areas, as well as through deaths resulting from cataclysmic events. As natural disasters increase in frequency and magnitude, so do the needs for population estimation to

accurately measure their impact, often requiring different data sources or implementation of new methods.

4. Combining airline traffic data with household surveys

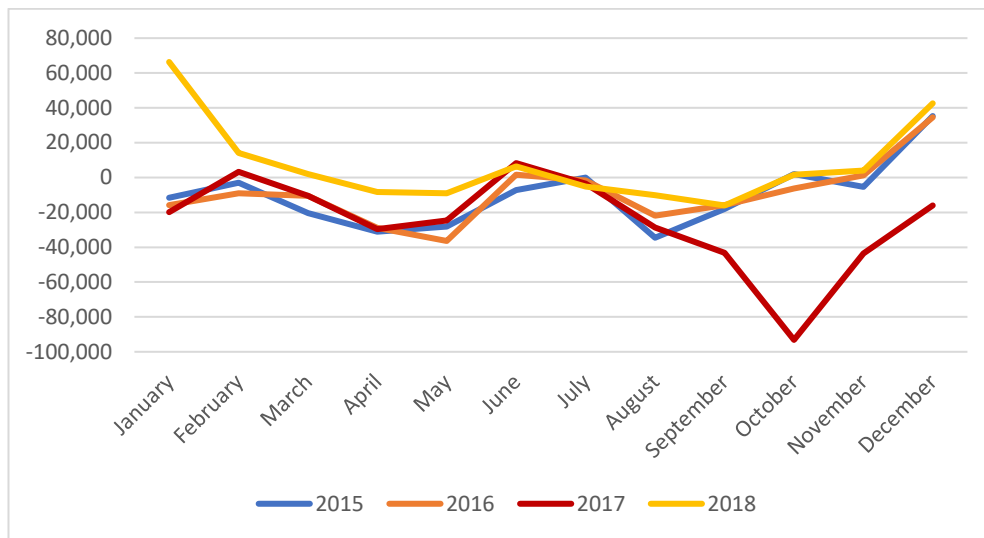
35. The United States Census Bureau initially used APT data, combined with the ACS and PRCS, to measure the impact of Hurricane Maria on migration to/from Puerto Rico. Household sample surveys like the ACS are not designed to pick up sudden mass movements of people, since retrospective survey-based migration data tend to “lag” actual migration events. Surveys do not measure a migration event in real time, but rather measure the event when the migrant is included in the sample. This works well when migration patterns are stable, but when there are large fluctuations in the magnitude of movement, these will not be fully picked up until later (usually in the following survey year). The late timing of Hurricane Maria (late September) in the 2017 ACS survey data collection cycle and the corresponding short period of time to be included in the sample created potential response complications. As a result, it was necessary to look for an alternative data source to measure the impact of Hurricane Maria on migration out of Puerto Rico.

36. In the methodology used after 2017, ACS/PRCS data were integrated (“blended”) with monthly APT data from BTS to improve previous estimates. Historically, APT data have consistently shown higher net out-migration from Puerto Rico to the United States than ACS/PRCS estimates. To account for this inherent difference between data sources, the revised method “blends” ACS/PRCS and APT data, which gave an estimate that better reflected the impact of Hurricane Maria.

5. Results

37. APT data showed a large net passenger outflow from Puerto Rico in the latter months of 2017 (September to December), followed by return flows in the early months of 2018 (Figure 1). This return movement to Puerto Rico from the United States dropped over the first quarter of 2018, returning to net outflow by April 2018. Before Hurricane Maria (20 September 2017), net movement between the United States and Puerto Rico followed relatively stable monthly patterns, with more passengers leaving than entering Puerto Rico, except for some summer or winter months (June and December, in particular). This corresponds with seasonal flight patterns, with more tourists coming during summer and winter months, as well as return visits by Puerto Ricans living in the United States during vacation periods.

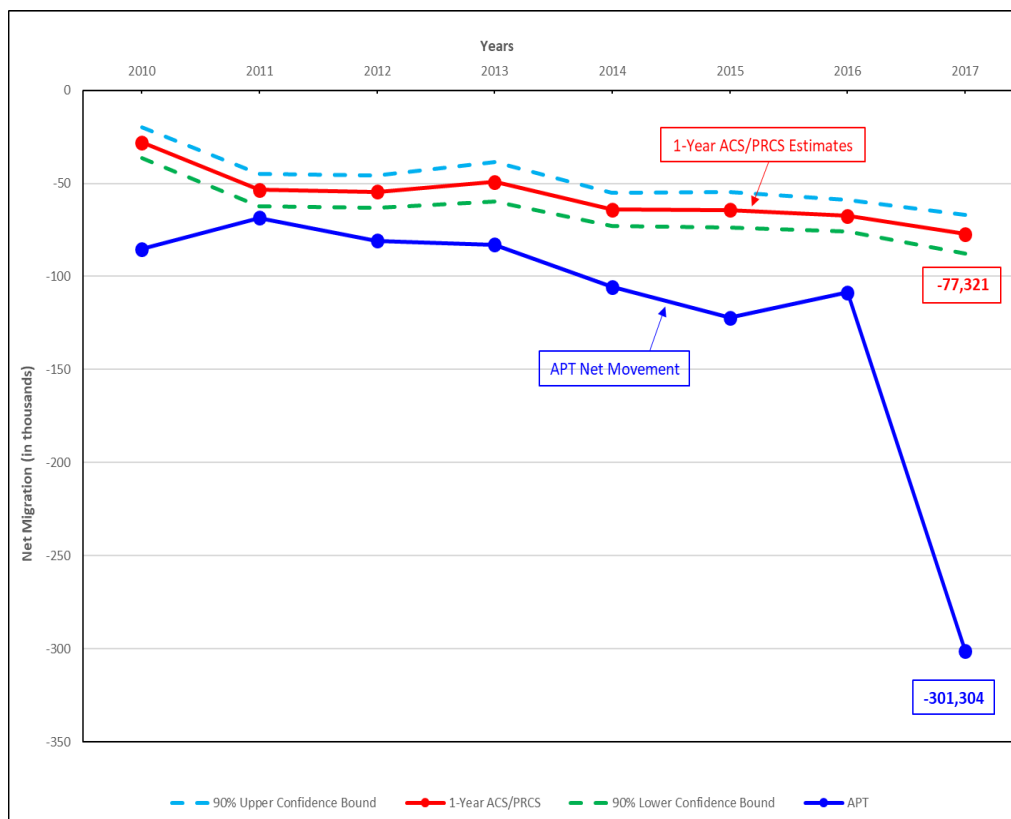
Figure 1
Puerto Rico monthly net flight passenger movements with the United States, 2015 to 2018



Source: US Bureau of Transportation Statistics Form 41, T100 (International) Segment All Carriers.

38. The “blended” methodology used for the 2018 estimates assumed that APT data were a better reflection of the true impact of Hurricane Maria on migration patterns than the ACS/PRCS, and that ACS and APT data followed similar historical patterns; thus, this relationship could be used to create a better estimate for the period. ACS and APT data tended to follow similar patterns before 2017, with APT data consistently showing more net out-movement from Puerto Rico than ACS net out-migration (Figure 2).

Figure 2
Puerto Rico net migration with the United States, 2010 to 2017

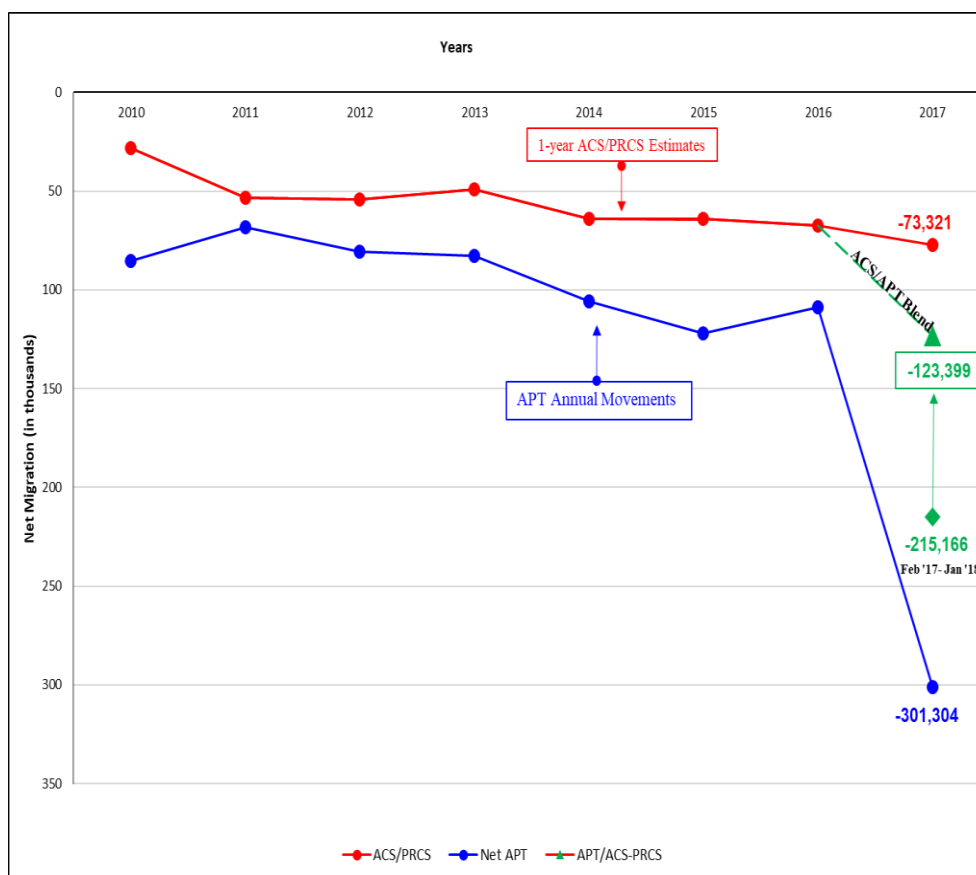


Sources: US Census Bureau, American Community Survey and Puerto Rico Community Survey; US Bureau of Transportation Statistics Form 41, T100 (International) Segment All Carriers.

39. Since the previous methodology had been based on the ACS/PRCS, efforts were made to make the two data sources as comparable as possible, compiling monthly flight data for the 2017 calendar year to coincide with the ACS/PRCS estimation period. Flight information was limited to domestic flights between the United States and Puerto Rico, excluding international flights. The method applied a simple ratio, using the ratio of ACS/PRCS-to-APT net migration results over the two years 2015 and 2016. Using ratios based on longer periods (as far back as 2012) was investigated, but a shorter period was chosen to reflect recent relationships between data sources more accurately. The calculated ratio was applied to the APT Puerto Rico-United States net migration figure measured for calendar year 2017, to remain methodologically consistent with previous ACS/PRCS-based estimates.

40. Since the Census Bureau's 2018 population estimates represent the population on 1 July 2018, return migration to Puerto Rico in early 2018 was also considered. To account for January return migration, the 12-month APT period from February 2017 to January 2018 was used, before applying an adjustment ratio to make the period as ACS-equivalent as possible, while still considering post-Hurricane Maria return migration to Puerto Rico. Shifting the period one month helped account for return migration, yet also kept most months (11 of 12) within the ACS/PRCS-equivalent 2017 calendar year. This modification resulted in an APT-Puerto Rico-United States net migration figure of -215,166, which when adjusted by the APT-ACS ratio yielded a final figure of -123,399 net migration between Puerto Rico and the United States (Figure 3). This "blended" method was again applied for the 2019 estimate. This was necessitated by the need to measure additional return migration to Puerto Rico in the aftermath of Hurricane Maria, which was not adequately measured in the 2018 ACS/PRCS. This resulted in a slight positive net migration gain for Puerto Rico, which had not been seen in previous decades.

Figure 3
Adjusted Puerto Rico net migration with the United States, 2010 to 2017



Sources: US Census Bureau, American Community Survey and Puerto Rico Community Survey; US Bureau of Transportation Statistics Form 41, T100 (International) Segment All Carriers.

6. COVID-19 adjustment

41. The COVID-19 pandemic greatly impacted movement to and from the United States starting in March 2020. While this had a dramatic impact on international migration flows, it also impacted movement to and from Puerto Rico, particularly towards the beginning of the pandemic. It was anticipated that the 2019 ACS/PRCS (which did not cover the pandemic period) would not be an adequate measure of 2020 Puerto Rico net migration patterns, so APT data was again used to make an adjustment. However, since the start of the pandemic only covered the last third of the estimates year (March–June 2020), the same blended method could not be used as it is based on a full calendar year of APT and ACS/PRCS data. As such, monthly flight data to/from Puerto Rico and the United States were used to make an adjustment to Puerto Rico net migration for the March–June 2020 period.

42. Flight patterns to and from Puerto Rico typically follow consistent patterns (Figure 4⁴). Typically, March through May are net passenger outflow months, while June, the start of the tourist season, is a passenger inflow month. The March 2020 APT data showed a large increase in net passengers to the mainland United States (out of Puerto Rico); April and May showed small positive net gains to Puerto Rico as the number of flights were drastically reduced; and June showed a large positive net gain to Puerto Rico as flights to and from the island increased.

⁴ 2017 and 2018 are excluded due to monthly anomalies created by Hurricane Maria.

Figure 4
Puerto Rico monthly net flight passenger movements with the United States, 2016, 2019 and 2020



Source: US Bureau of Transportation Statistics Form 41, T100 (International) Segment All Carriers.

43. To calculate the 2020 COVID adjustment factor, using APT data, the net passenger total was calculated for the months of March to June 2020. Next, a ratio was calculated using the historical average net passenger movement for March to June 2010–2019, excluding 2018 (Hurricane Maria), and the March to June 2020 net passenger total. The ACS/PRCS seasonal total was then reduced by multiplying one-third of the 2019 ACS/PRCS annual total by the net passenger movement ratio. For the final estimate, two-thirds of the 2019 ACS/PRCS estimates were applied to the ACS/PRCS reduced seasonal total. The overall impact of the COVID-19 adjustment was reduced net out-migration from Puerto Rico.

7. 2021 air passenger traffic-based method

44. Starting in 2021, the United States Census Bureau used APT data to directly measure net migration to and from Puerto Rico and the rest of the world, although the ACS/PRCS will still be required to estimate inflows, outflows and the demographic characteristics of movers. In recent years, several major events have impacted migration patterns to and from Puerto Rico and necessitated methods to adjust survey-based estimates with flight data. These events have included Hurricane Maria in September 2017, a large earthquake in southwest Puerto Rico in January 2020, and the COVID-19 pandemic beginning in March 2020. Moving to a flight-based method to compute net migration will improve the accuracy and recency of net migration estimates for Puerto Rico and reduce the number of future adjustments needed to account for major events impacting migration.

45. There are several inherent limitations with the former methodology that are resolved with a flight-based methodology:

(a) Given that survey data are not produced instantaneously, the timeframe from when the data are collected, processed, and produced is different from the timeframe that is being estimated. Thus, changes in the population are not apparent until the following estimate year;

(b) The methodology only included movement from Puerto Rico to/from the United States with an inaccurate assumption of net zero migration to/from United States Island Areas and foreign countries. The PRCS can measure in-migrants to Puerto Rico from outside the United States but cannot measure out-migrants to abroad. Assuming net zero

migration between these locations is not ideal due to Puerto Rico's proximity to the Virgin Islands, the Dominican Republic, Haiti, and Cuba;

(c) The period covered by the ACS/PRCS (January–December calendar year) does not align with the estimates year (July–June);

(d) Since 2018, data overrides and projections have been necessary to maintain data accuracy and reliability, which will be less likely in the future;

(e) The ACS/PRCS will soon be impacted by differential privacy. Differential privacy, a method of disclosure avoidance modernisation adopted by the Census Bureau, is designed to increase security of individual level data by injecting noise into data to minimise precision and record vulnerability⁵. The accuracy of ACS/PRCS data could potentially be impacted in the future, though to what extent is currently unknown. APT data are independent of Census Bureau data and therefore will not be affected by this change.

46. Annual net migration totals will be calculated using in/out passenger flow data for the estimate year (June–July) using APT data: in-bound total passengers to Puerto Rico are subtracted from out-bound total passengers from Puerto Rico to obtain the net migration estimate. This includes international passengers, unlike the “blended method” previously described. Typically, domestic flight data are received with a 3-month lag, while international data are available 6-months later. However, the Census Bureau has an internal agreement with BTS to access international data with only a 3-month lag, enabling the inclusion of current international flights in the estimates.

47. Since flight data only give a net flow estimate, and do not include any specific in- or out-migration flows or demographic characteristics, this information will continue to be obtained from the ACS/PRCS. Sex will be tabulated using the one-year PRCS (ROYA) data for migration inflows, and then distributed to single years of age using within-sex proportions from PRCS (ROYA) five-year estimates. This process will be repeated for migration outflows using the ACS (ROYA) one-year estimates for sex and the ACS (ROYA) five-year estimates for age proportions within sex.

48. As previously noted, seasonal variations related to tourism, particularly during the summer and winter months, may lead to trends that require annual fluctuation adjustments. The fluctuation adjustments will be dependent on the measurement period to balance the seasonal inflow and outflow trends.

III. Reasons for not using new data sources

49. Based on the results of the survey, only a small number of countries in the UNECE region currently use or are considering using new data sources, such as big and social media data, for measuring migration and cross-border mobility. Out of 36 countries that responded to the survey, only 5 countries reported on current activities/projects that use new data sources, and 4 more countries are exploring this possibility. So, in total about one quarter of the countries use – or is considering the possible use of – new data sources for measuring migration and cross-border mobility.

50. Why are most countries not using new data sources for measuring migration and mobility? Do they use new data sources in other statistical fields? What factors can influence – positively or negatively – the use of new data sources for measuring migration and mobility?

A. Use of new data sources for producing statistics in other statistical fields

51. The countries that currently are not using new data sources for measuring migration and cross-border mobility were asked in the survey whether new data sources are used by the

⁵ Hawes, M. (2021) Understanding the 2020 Census Disclosure Avoidance System: Differential Privacy 101 Webinar. US Census Bureau. <https://www2.census.gov/about/training-workshops/2021/2021-05-04-das-presentation.pdf>.

NSO for producing statistics in other statistical fields. Out of the 24 countries responding to this question, 8 countries (one third) reported using – or testing the use of – new data sources in other statistical fields, while the others replied “no” (9 countries) or “don’t know” (7 countries).

52. France and New Zealand reported the use of new data sources to produce statistics on internal migration. In New Zealand, mobile phone data are combined with administrative data on border-crossings to estimate the geographic distribution of international visitors within New Zealand. In addition, integrated administrative data from across a variety of government agencies is used to estimate internal migration. Integrated administrative data is also used to supplement and validate census information, validate official estimates of resident population, and estimate international migration by ethnicity.

53. Four countries reported the use of new data sources to produce statistics on internal mobility and commuting. In Portugal, Facebook's "Data for Good" Initiative is used to produce population mobility indicators at regional level. In Spain, experimental statistics were produced on commuting through mobile phone data. Switzerland reported a pilot project for collecting data on daily mobility via smartphones: in the framework of a survey, a sample of the participants install an app on their smartphone that collects data on their daily mobility – instead of reporting their moves in a questionnaire. In the United Kingdom, the Data science campus⁶ uses various new data sources, including exploring the use of Facebook data for movements within the country.

54. Six countries reported experiences on the use of new data sources in different areas of economic statistics. In the area of price statistics, Portugal uses web-based data, Lithuania data from supermarkets, and Slovakia both web-scraping and scanner data of retail chains. On tourism statistics, Slovakia uses data from websites of accommodation booking systems, and Spain did some research on the use of credit cards for the estimation of tourism expenditures. On international trade, Portugal reported studies on using internet searches from specific sites to validate statistics on international trade, including on volumes and prices of international trade in electric energy. Portugal also uses the E-invoice system (E-factura) from Tax and Customs Authority for statistics on regional economic activity, and internet searches for online job vacancies and enterprises sites, to produce labour market statistics. Finally, France uses new data sources for statistics on consumption, and the United Kingdom for timelier indicators of the economy.

B. Reasons for not using new data sources

55. The 24 countries that do not use new data sources for measuring migration and cross-border mobility were asked in the survey what were the main reasons for that. The survey offered several possible reasons, in addition to allowing a write-in response, and respondents could select up to three reasons (Table 1).

Table 1

Reasons for not using new data sources for international migration statistics

<i>Reasons</i>	<i>Countries (out of 24)</i>
Individual privacy concerns, legal constraints and rules that limit the use of big data and new sources for statistical purposes	12
Already produce high quality statistics about migration through traditional data sources	11
Lack of specialised staff in managing big data	8
Big data are not suitable for official statistical use	6
Lack of agreement among agencies (e.g. NSO and line ministry) and potential providers	3
Lack of funds to access big data	1

⁶ <https://datasciencecampus.ons.gov.uk/>

<i>Reasons</i>	<i>Countries (out of 24)</i>
Comments and/or other reasons	3

56. The reason mentioned by most countries was individual privacy concerns, legal constraints and rules that limit the use of big data and new sources for statistical purposes (12 countries).

57. Eleven countries replied that they already produce high quality statistics about migration through traditional data sources. Several of those countries specified that they use administrative data, and that they consider that new data sources would not give much added value. New Zealand specified that since 2019 (with estimates backdated to 2001) it uses administrative data on travellers crossing the border to estimate international migration with high accuracy and timeliness. New Zealand added that obtaining commercial/non-governmental data sources in a robust and continuous way is generally challenging and does not offer much value above what can be available through an integrated administrative data system.

58. Several countries cited the lack of specialised staff in managing big data (8 countries) and the consideration that big data are not suitable for official statistical use (6 countries). Some countries provided additional information on the reasons why they consider big data as not suitable for official statistical use. In Slovakia, population and migration statistics are produced based on administrative data, and in addition there would be no legislative support on the use of new data sources. Similarly, in Switzerland, data on international migration flows or the migrant population come mainly from administrative data or survey data that are already available at low cost and of good quality. These data are used for public management and policymaking in the field of migration, and there is no need for more extensive data on population movements. Moreover, Swiss official statistics are mainly focused on the population that is legally and permanently residing in the country. It therefore does not consider populations that are not represented in official registers, such as undocumented migrants or short-term migrants, for whom the use of new data might be more relevant. In Spain, INE worked with mobile phone operators to produce data on commuting (and also, to a lesser extent, on tourism) but no valuable information was produced for migration statistics, and no other suitable big data source was identified. The United Kingdom noted that more assurance is needed on what the new data sources can give for migration statistics. Before using these data in the production of statistics, the data's ability and accuracy to report on migration flows, distinctly from other movements of people, should be verified. In general, with big data it can be difficult to identify migrants or migration within the vast amount of data. For example, with mobile phone data, it is difficult to disentangle movements across borders that are not just visits or short trips.

59. The lack of agreement among agencies (e.g. NSO and line ministry) and potential providers was selected as a reason by three countries. In the Czech Republic, migration data are collected from different administrative sources, including the Ministry of the Interior (Foreign Police Service) which is one of the main providers. It was suggested that any initiative to use big data should come from that Ministry, as it would require their modification of the Foreign Information System.

60. The lack of funds to access big data does not seem to be a major obstacle, as it was cited by only one country.

IV. Significant innovations on measuring international migration

61. As detailed in the previous sections of this report, most NSOs that participated in the survey (on the use of alternative data sources in migration statistics) indicated that big data was not currently being used.

62. However, 12 NSOs indicated that they were “working on significant innovations on measuring international migration”, including “new approaches using current sources and/or exploring new data sources”. This section summarises the findings from these responses, split into two themes:

- (a) Changes in data sources from what was being previously used by the NSO;
- (b) New or innovative methodology that NSOs are using on this data, or on their existing data.

A. Data-related innovations in measuring international migration

63. Most countries that are exploring data-related innovations were starting to use, extending the use of, or enhancing current estimates using administrative data (also referred to as admin data). Of the respondents mentioned above, eight indicated that they were engaging with administrative data in some way, with different levels of maturity and progress along this path.

64. Three countries indicated that they were using surveying to estimate migration, obtain demographic breakdowns of migration, or for population estimation. In some cases, these were supplementing administrative data driven measures.

1. Summary of administrative data use

65. Countries are at different levels of maturity in their use of administrative data (Table 2). This section uses country examples based on detailed responses to the survey to highlight different stages of development and the types of approaches that are being explored in the administrative data space.

Table 2

Summary of use and exploration of administrative data for estimating international migration across respondent countries

<i>Maturity of administrative data use</i>	<i>Country</i>	<i>Details of use</i>
Exploration/early development	France	A project is underway to use and merge administrative data for estimating migration.
	Portugal	Exploring administrative data for annual resident population estimates and achieve coherence between migration flows and migrant stocks.
	Switzerland	Has very mature and embedded usage of administrative data in the production of statistics. Currently developing the production of longitudinal data on international migration flows based on their already mature register data.
Mature development/pre-production	Armenia	Use of population register with border management data is planned for use in 2022 Census. After census, these sources may be used for migration measurement.
	Hungary	Improving migration statistics based on multiple (unspecified) administrative data sources.
	Slovakia	2021 Census based on integration of administrative data with additional migration data from statistical surveys.
	United Kingdom	Shifting from survey-based migration statistics to new measures based on administrative data.

<i>Maturity of administrative data use</i>	<i>Country</i>	<i>Details of use</i>
	United States	Use of use complete flight data to estimate net migration for the Commonwealth of Puerto Rico, supplemented by demographic attributes from survey data.
In production	New Zealand	Administrative data is used as the main source for migration estimates, with statistical models used to give early provisional estimates.

2. Selected examples on innovations with administrative data

A. Portugal

66. Statistics Portugal is studying and exploring the use of administrative data in the production of official population statistics, based on a variety of sources including:

- (a) National population civil register;
- (b) Foreign population register;
- (c) Education attainment register;
- (d) Tax register;
- (e) Social security register;
- (f) Employment and unemployment registers.

67. These records are being used to construct a resident population database, to be updated annually. In transitioning to the use of administrative data, coherence between migration flows and migrant stocks is a key concern. Comparing the resident population between reference dates can be used to measure migrant stocks, but this does not give the gross migration flows between the reference dates. Statistics Portugal is currently studying methods to overcome the challenges and limitations of the administrative data sources to measure migration flows.

B. Armenia

68. Armenia is planning to use a Population Register in combination with data from their Border Management Information System, for the 2022 Population Census. For the first time, these administrative data will be combined with a large sample survey (involving 25% of the addresses in the population register). This census approach has the potential to give new measures of migration and mobility. The systems for the combining of administrative data and the survey are in their final stage of development.

C. United Kingdom

69. The Office for National Statistics (ONS) is moving towards an administrative data-based measure of migration, replacing the measures based on the International Passenger Survey (IPS), which had long-recognised limitations. This has been further accelerated by the emergence of the COVID-19 pandemic, which caused the IPS to be suspended by the ONS in 2020, although it has now resumed, focusing on international travellers. The ONS is using a variety of administrative sources, mainly integrated through their Registration and Population Interaction Database (RAPID)⁷ developed by the Department for Work and Pensions (DWP). This gives a single coherent view of interactions across the systems in DWP, HM Revenue and Customs (HMRC) and local authorities via Housing Benefits. These

⁷ RAPID contains a record of everyone who has a National Insurance number. For more detail, see [International migration: developing our approach for producing admin-based migration estimates: Measuring international migration with administrative data.](#)

interactions include income support benefits, employment, self-employment, pensions and in-work benefit.

70. Adjustments are made to account for the variety of migrant types that have differing levels of representation across the administrative system. To overcome the lags associated with classifying migration based on this data, the ONS also undertakes modelling to estimate migration flows, discussed in the section 2 of the document.

3. Summary of survey data use

71. Table 3 summarises how survey data is used or explored by the countries who responded.

Table 3

Summary of use and exploration of survey data across respondent countries, by estimated stage of maturity in development and use of administrative data, based on the responses received

<i>Maturity of survey data use</i>	<i>Country</i>	<i>Details of use</i>
Mature development/pre-production	Armenia	A 25% sample survey will be used to supplement the administrative data based on 2022 Census. (See earlier section Armenia)
	Czech Republic	Migration data about from Labour Force Survey, including date of migration and place of birth of parents.
	United States	Survey provides age-sex attributes of migrant flows for the Commonwealth of Puerto Rico.

B. Methodology-related innovations in measuring international migration

72. Some countries also indicated details about methodological developments, independent of data driven innovations, that they were undertaking. Elaboration was given by New Zealand and the United Kingdom, and they will be used as brief examples of different methodological innovations that are taking place.

1. Selected examples of methodological innovation

A. New Zealand

73. New Zealand collects arrival and departure information, based on travellers' passports and flights for all passengers. Using this data, it has developed an outcomes-based measure of migration, which classifies immigrants and emigrants based on the amount of time a traveller spends in or out of the country⁸. In general, this requires observations of a travellers' border-crossings for up to 16 months, which necessarily introduces a delay in producing timely finalised estimates of migration. Therefore, Statistics New Zealand uses a predictive machine learning to give provisional estimates of migration, based on predictively classifying border-crossings.

74. The model learns about the features of border crossings that make them more or less likely to be migrant crossings, by looking at millions of historical records. The model looks at "features" in the historical border-crossing data, such as the direction of crossing, and date of the crossing, the amount of time in/out of country, the amount of time that has passed since the border crossing. It also learns about attributes from the passport data, such as the age, sex, and citizenship of travellers, and the broad type of visa they hold. Each of the records also has information on whether a particular crossing was a migrant crossing or not (the outcome).

⁸ Details of the specific classification rules are available at [Migration Data Transformation: The 12/16-month rule](#).

This allows the model to establish links between different combinations of the features above, and the likelihood that a crossing is a migrant crossing. The model uses a technique called gradient boosting to do the learning. This is implemented using a well-established algorithm called “XGBoost”.

75. This process creates an ensemble of predictors that can be applied to border crossings where the outcome is unknown, based on the features in unknown border crossings. Using the established probabilities, the number of migrant crossings for a particular reference period, aggregated by different attributes, including age, sex, and country of previous residence are estimated.

B. *United Kingdom*

76. As briefly discussed above, the Office for National Statistics (ONS) is moving to administrative data for its measurement of migration. Migration is estimated based on indicators of activity within the data. In general, arrivals with 12 months of activity are classified as migrants, although some exceptions and adjustments are required (details are available in footnote **Error! Bookmark not defined.** above). In addition to the adjustments for coverage and other factors, the ONS is also developing modelling methodology to increase the timeliness of migration estimates, given that, by the current standard definition, a minimum of 12 months must pass for classifications to be available through observing activity in the administrative data for 12 months⁹. Although the ONS are exploring the use of a broader set of definitions to sit alongside the UN metric. These will aim to provide real-time measures of the population and address increasing user demand for these measures that has intensified during the pandemic.

77. The United Kingdom is exploring the use of state-space models to complement the administrative data driven approach, which may be able to give provisional predicted migration. These models have so far been used to estimate migration up to December 2020¹⁰. The intention, as presented in literature published by the ONS, is to explore the use of such models to improve the timeliness of the administrative data driven migration estimates discussed previously¹¹.

V. A literature reference tool: Database on Innovations in Migration Statistics (DIMiS)

78. There are several papers on migration using innovative data sources, published in the scientific journals or as working papers, of potential interest for official statisticians. In order to bridge the gap between research and practice, the Task Force selected some of those papers and, with the support of the UNECE Secretariat, developed a database¹² containing references of published works in migration-related matters using innovative methods and/or new data sources. This *Database on Innovations in Migration Statistics (DIMiS)* is meant to raise the awareness about the scientific progress in this domain and to make it easier the access to relevant information for official statisticians and general practitioners interested to the subject. DIMiS aims to cover studies using ‘non-traditional’ data sources as well as innovative methods applied to ‘traditional’ data sources, relevant to producing statistics on migration (in its broader meaning).

79. For this initial version of the database, about one hundred papers have been identified, covering topics such as international migration, internal migration, human mobility,

⁹ Details on the timeliness of the methods and latest developments are available at [International migration statistical design progress report](#).

¹⁰ Main findings from the latest modelling estimates are presented in [Long-term international migration, provisional: year ending December 2020](#).

¹¹ Discussion of the implications and limitations of the modelling, and future use, is presented in [Using statistical modelling to estimate UK international migration: Discussion](#).

¹² A review of selected scientific works is done in the European Commission Joint Research Centre (JRC) Science for Policy Report on “*Data Innovation in Demography, Migration and Human Mobility – Data gaps, opportunities, and challenges of non-traditional data*” by Iacus, S.M., C. Bosco, S. Grubanov-Boskovic, U. Minora, F. Sermi and S. Spyrtatos (forthcoming).

population displacement, and population distribution. Of those papers, only those applying to the migration stocks and flows (either international or internal) have been included in this first list. Depending upon the interest of the users, the range of topics can be enlarged or further disaggregated (e.g. distinguishing between flows and stocks, or specific for asylum seekers and refugees, integration of migrants, etc) in future versions of DIMiS.

80. DIMiS is structured according to a set of key variables that can be used as filter(s):

(a) *Author(s)*: names of the author(s). The first author's name is presented with the surname first, followed by the initial(s) of the first and middle names. For all the other authors, first the initial of the forename and then the full surname is listed. The list of authors is reported entirely and thus not abbreviated with "*et al.*" It is therefore possible either to search for a specific surname, or to sort the articles by first author's name;

(b) *Year*: year of publication. When the online publication year is different from the printed publication year, it is preferred the most recent year, as this may correspond to a version including corrigendum. For non-peer-reviewed articles, for instance such as those in SocArXiv (the open archive of the social sciences), it is taken the year of first uploading/dissemination, as the paper may undergo several revisions over time, also after its inclusion in DIMiS;

(c) *Title*: title of the article. No uniform style as for the use of capital letters, which mimics the one adopted in the articles;

(d) *Journal*: scientific journal/book/research project/website repository that contains the paper. No uniform style applied;

(e) *Abstract*: the published summary of the purpose, method, and outcome of the article – when available;

(f) *Topic(s)*: the main migration-related aspect discussed in the article/paper. The following categories have been considered for this initial version of the database:

- (i) International migration;
- (ii) Internal migration¹³;
- (iii) Human mobility;
- (iv) Population displacement;
- (v) Population mapping/distribution;
- (vi) Other.

Only the papers whose topic covers the first two categories above are included for now in DIMiS, even though they may cover other topics as well. Multiple topics covered in a single paper are listed sequentially in the field, in accordance with the list above. Therefore, it will not be possible to find papers that are exclusively in any category(-ies) other than a) or b), but it is possible to find papers covering also categories c) and/or further.

(g) *Geography*: the country(-ies) covered by the study, with the pertinent sub-national geographic area(s) indicated between parentheses. In the case of purely methodological papers or for review articles exploring the state of play of the scientific literature, this key does not apply and the field is filled with "not applicable".

(h) *Data Sources*: the source of the data used in the article. The following categories have been considered:

- (i) Mobile Network Operator(s) (MNO);
- (ii) Social media (e.g. Twitter, Facebook, WhatsApp, LinkedIn);
- (iii) Search engines (e.g. Google);

¹³ For the sake of this report, migration between USA and Puerto Rico is classified as internal migration.

- (iv) Web demographics (databases of personal microdata that are available on the World Wide Web);
- (v) Registers (administrative data);
- (vi) Survey (e.g. census, household sample);
- (vii) Others (such as emails, internet providers, traffic data, satellite images, specific databases).

In general, a data source has its own category when there are at least two articles using that. An article may exploit one or more *data* sources, which are all reported in this field. To the possible extent, the specific data source is reported between parenthesis (e.g. Facebook within the ‘Social media’ category). Users can then run a query searching for all papers using a category of data source (e.g. MNO), or a specific one (e.g. Twitter).

81. Of the 52 papers included in this initial version of DIMiS, 36 articles are on international migration (alone or covering also other topics), 23 on internal migration (alone or covering also other topics), and 7 papers cover both international and internal migration (see Table 4).

Table 4

Articles in DIMiS by topic

<i>Topic</i>	<i>Number of articles</i>
Internal migration	9
Internal migration / human mobility	4
Internal migration / population displacement	3
International migration	25
International migration / human mobility	1
International migration / internal migration	3
International migration / internal migration / human mobility	4
International migration / population displacement	3
Total	52

VI. Conclusions and next steps

82. Experiences in the NSOs are *limited in the use* of big data and the material collected for this report might not be sufficient to provide specific recommendations. However, the *potential of new data sources* to address both persistent and emerging migration issues and urgent data needs is promising. Sharing and updating of relevant *information and tools facilitate the use* of big data and new data sources for measuring migration and cross-border mobility.

A. Limited use so far

83. Many countries face limitations on the use of big data; in particular, many countries reported individual privacy concerns, legal constraints and rules as the reason for not using new data sources. To address this, it is important to consider ethical and privacy issues in the international debate. At the same time, it is important to use methodologies and techniques that maintain the confidentiality of information while enabling statisticians and researchers to utilise the potential of big data.

84. Many countries currently prefer focusing their efforts to the use of administrative data and the integration of data coming from different sources, which may be related to difficulties faced in accessing big data. Considering administrative data, what is a new data source for one country is sometimes a traditional source for another country. It depends on many different aspects such as the specific history of the countries, the legal and institutional frameworks and the available resources.

85. Even in countries that reported several experiences on the use of big data in other statistical sectors (tourism, commuting, economics, etc), the exploitation of big data in the field of measuring migration is still limited. There are many reasons for this. Many countries have a long and shared history of migration statistics and have often shared using international standards and definitions with international organizations. According to the survey, these countries tend to prefer to strengthen the quality of traditional data rather than explore new data sources which require overcoming multiple administrative and legal obstacles.

B. Potential benefit

86. The analysis carried out by the Task Force confirms the potential of new data sources for measuring migration, including emerging patterns of migration and cross-border mobility, and for addressing urgent data needs arising from unexpected natural events (pandemics, hurricanes, etc) and geo-political events (wars, persecutions, political crises, etc).

87. Big data and new data sources could help bridge the gap between the need for timely information and the time required by the official statistical collection process. They could therefore play an important role in improving the quality of currently produced statistics. Many countries are moving towards integrated statistical systems, and new data sources could be important resources to be integrated into these systems. It is therefore hoped that in new integrated statistical systems for measuring migration, big data and new data sources will find an appropriate place and can enhance the quality of migration statistics.

C. Next steps – maintaining the Database of Innovation in Migration Statistics

88. This report shows ongoing work on the use of big data and new data sources in NSOs, which are expected to be useful for countries interested to work in this area.

89. The Task Force created the Database of Innovation in Migration Statistics (DIMiS) to expand the perspective by including research and studies about migration and big data conducted outside NSOs, such as articles and papers dealing with big data for migration statistics (including international migration, human mobility, internal migration, etc). The Database facilitates exchange of information about projects and best practices to encourage NSOs to consider research experiences around the world in different fields of migration studies.

90. NSOs and researchers are invited to report any new experiences and publications to UNECE so that the Database can be regularly updated. DIMiS is not an exhaustive repository, but it is intended as a first step towards a living repository that should be constantly fed with new experiences to facilitate the sharing of projects based on new data sources.

91. The Database is intended to be flexible. The information included could vary and increase as the experiences become broader and more complex, allowing DIMiS to become a key reference supporting NSOs in the use of new data sources and related method.