

Chapter 27

Migration Statistics



Albert Kraler and David Reichel

As a first point of reference, migration statistics—that is, general indicators on the size, composition, and characteristics of the migrant population and of migration flows—are ubiquitous and a standard element of both official population statistics and academic accounts of migration dynamics. What is more, statistical indicators on migration have become part of the “steady diet of statistics” (Starr & William, 1987) that the public is served through the [media](#) and other channels by a variety of actors, including [public institutions](#), [intergovernmental organisations](#), [NGOs](#), politicians, and activists.

Numbers have thus become an ever more integral part of academic, political, and broader societal discourses on migration and thus also of its construction as a societal ‘problem’. Indeed, as Bill Gates has reportedly advised Australian billionaire Andrew Forrest, founder and funder of the Global Slavery Index, “if you can’t measure [a problem], it doesn’t exist” (Behrmann, 2013). Yet, statistics not only serve as neutral evidence of the existence of a phenomenon and its quantity, but they also have the power to push specific dimensions of a phenomenon into the limelight and so the potential to change both public perceptions and policies. In this context, statistical indicators often turn into benchmarks of the success or failure of particular policies in a given area (see more broadly Boswell, 2018).

Statistics on border apprehensions or migrant fatalities—largely absent from the public realm 20 years ago—are a case in point. Now delivered on an almost daily basis, such statistics are—depending on one’s political position—used to decry the

David Reichel works for the European Union Agency for Fundamental Rights (FRA). The views expressed in this article are solely those of the authors and its content does not necessarily represent the views or position of the FRA.

A. Kraler (✉) · D. Reichel
European Union Agency for Fundamental Rights (FRA), Danube University Krems, Krems,
Austria
e-mail: albert.kraler@donau-uni.ac.at

deadly consequences of European border politics, or conversely to denounce the failure of these policies to protect European borders—or migrants from criminal smugglers (Heller & Pécou, 2020). The creation of integration monitoring systems is similarly deeply intertwined with the emergence of integration policy as a separate policy field (see Bijl & Verweij, 2012). In the context of integration policy, the regular production of indicators on migrant integration has been argued to be a precondition for effective policymaking, both on the national and the EU level. At the global level, most policy initiatives of the past two decades in the area of migration and beyond—the 2015 Sustainable Development Goals of which 11 make reference to migration and mobility, and more specifically, the Global Compact on Migration and the Global Compact on Refugees—follow a similar line of reasoning and contain specific provisions on data and evidence needed to monitor ‘progress’ (Ahmad-Yar & Bircan, 2021). As Anne Gallagher has observed in the context of data on trafficking and ‘modern slavery’, “[t]his is part of a much broader trend [of presuming] quantification of a practice or phenomenon identified as problematic [as] an essential pre-requisite to any kind of meaningful action.” (Gallagher, 2017; see more broadly Andreas & Greenhill, 2010).

At the same time, statistical indicators on migration and the underlying categories they refer to are not self-evident and are, at times, contested. Moreover, they are subject to change over time, both in terms of the categories used for measurement (e.g. ‘migrant’, ‘foreigner’, ‘foreign worker’, etc.) and in terms of the specific definitions used to measure those categories.

This chapter provides a critical appraisal of data collection and statistical data production on migrants and migration, indicators used to identify migrants, and the different uses of migration statistics for political and scientific purposes. The chapter proceeds as follows: We will first review the history of data collection on migration and consider international efforts to harmonise migration statistics on the international level. We will then provide an overview of different ways to measure migration, core concepts, and definitions used to measure migration. In the subsequent section, we will describe different data collection methods and sources used to produce statistics on migration, including possible ‘new’ data sources for measuring migration, often referred to as ‘big data’. In the last section, we will briefly demonstrate different ways of analysing aggregate migration data.

27.1 A Brief History of Migration Statistics¹

27.1.1 *Migration Statistics*

Statistical data collection on migration has emerged simultaneously with the emergence of other modern practices of registration and counting developed to render societies ‘legible’ (Scott, 1998; Torpey, 2000). These in turn were embedded in

¹This section is adapted from Kraler et al. (2015).

broader dynamics of standardisation and measurement linked to the expansion of capitalism in the modern era from about the fifteenth to seventeenth centuries. As Schrover (Chap. 2 in this volume) shows, there are thus—for non-historians perhaps surprisingly—extensive historical statistical data sources on migration, even if these are patchy and geographically biased, in terms of who was measured and what categories were used to describe migrants.

Efforts to harmonise international data collection on migration go back to the nineteenth century. The first major international discussions on migration statistics were held at the 1891 congress of the International Statistical Institute in Vienna, and subsequent conferences in 1901 and 1903 (United Nations Department for Social Affairs, 1949). The first attempts to standardise definitions in migration statistics followed 30 years later at the International Conference on Emigration and Immigration held in Rome in 1924. When the International Labour Organization (ILO) was created in 1919, it was also mandated to support the international movement of people. It started its own international data collection on international migration in the early 1920s (see Schrover (Chap. 2), this volume), and subsequently became the main actor in regard to promoting efforts to improve international data collection on migration during the interwar period.

An ILO-sponsored conference on migration statistics in 1932 adopted the first more systematic set of recommendations for the improvement of migration statistics (United Nations Department for Social Affairs, 1949, p. 1). From the outset, achieving comparable international migration statistics was considered important not only for statistical, or for that matter, scientific, purposes; it was also seen as a precondition for “the regulation of migration by international convention [and a tool to] facilitate cooperation of the administrative authorities of different countries” (International Labour Office, 1932, p. 86, quoted in Kraly & Gnanasekaran, 1987, p. 968). The 1932 recommendations were revised in 1953 and 1976, but both times not widely implemented. The latest revision dates from 1998 (Herm, 2006; United Nations, 1998). It was preceded by several conferences and studies, also [drawing on expertise](#) from outside the UN system such as the Council of Europe, the OECD, IOM, and Eurostat, all of which had become key players in policies on statistical data collection by the time of the 1997 revision (Herm, 2006). At the time of writing, a further revision is in preparation to: (1) address new patterns of migration presenting challenges to statistical measurement, and (2) to better align data collection with policy needs (UN Statistics Division, 2018).

27.1.2 *Refugee Statistics*

Statistics on [refugees](#) and other humanitarian migrants have largely been excluded from these efforts to develop harmonised concepts and definitions, even if there was a recognition that refugees were—at times—an important component of population movements across borders. Ultimately, refugees were thus defined by the legal or institutional regime they fell under and their counting was left to the respective

institutions in charge of a particular legal or institutional regime. In the context of migration statistics, however, statistics on refugees tended to be ignored. Reflecting an absence of a shared understanding about the essence of the concept of refugees at the eve of UNHCR's creation, the UN Population Commission, reporting in 1949, noted the importance of refugees. However, it refrained from providing a definition, only referring by way of example to "persons seeking entry without having the papers normally required for admission" as one category that could fall under a statistical refugee definition (quoted after United Nations Department for Social Affairs of November, 1949, p. 14). Today, of course, such persons would squarely be regarded as *irregular migrants*. Subsequently, the refugee definition of the Geneva Refugee Convention became the main basis for statistical reporting on refugees (UNHCR, 2020), and UNHCR the main body collecting data on refugees.

However, before the 1990s, UNHCR collected data only on a country basis and without any guidelines on how to collect them. Although collating such data occasionally for internal purposes, these data were rarely published. This changed only in the mid-1990s when the organisation as a whole took a more knowledge-oriented turn. From 1994 onwards, it started publishing statistical data more systematically and invested in improving its data collection. Already before that time, the US Committee for Refugees (USCR) regularly published an annual statistical report ("World Refugee Survey") on refugees in need of assistance and lacking opportunities for permanent settlement. This definition excluded most refugees in Europe, North America and Australasia, anticipating the concept of protracted refugee situations coined by UNHCR in the early 2000s (Crisp, 1999; Czaika & Kraler, 2020). Controversially at the time, the USCR also included Palestinian refugees under the United Nations Relief and Works Agency (UNRWA) in their refugee definition. The organisation was also the first to regularly report on Internally Displaced Persons. On a more systematic basis the Internal Displacement Monitoring Centre (IDMC), established by the Norwegian Refugee Council in 1998 in the context of the adoption of the Guiding Principles on Internal Displacement, took up the task of data collection on internal displacement and continues to be the key data source in this area.²

UNHCR too had started to report on IDPs (although more selectively) in the mid-1990s, subsuming it under the newly coined category "population of concern", alongside *asylum seekers*, returned refugees and returned IDPs and various other groups. In its 2009 *Global Trends* report (UNHCR, 2010), UNHCR reframed populations of concern to "forcibly displaced", adding Palestinian refugees under UNRWA to the categories counted. By the 2014 edition, UNHCR's report was published under the new title *Global Trends: Forced Displacement in 2014* (UNHCR, 2015), completing a journey of conceptual expansion that started off from statistical reporting driven by a purely administrative-institutional logic catering for specific categories that UNHCR was mandated to care for, and ending that journey with the implicit claim to present an authoritative account of all global

²See <https://www.internal-displacement.org/about-us>.

forced displacement. At the same time, UNHCR was well-aware of the flaws of international refugee statistics, and together with Statistics Norway proposed the organisation of a conference on refugee statistics and the development of recommendations on refugee statistics to the UN Statistical Commission in 2014. At the conference held in October 2015 in Turkey, an Expert Group on Refugees and Internally Displaced Persons was appointed, whose recommendations were released in 2018 (Eurostat/Expert Group on Refugee and Internally Displaced Persons, 2018). In parallel, it also oversaw a report on statistics on internally displaced persons (Eurostat, 2018).

27.1.3 Migration Statistics in the European Union

In the European Union context, efforts at collecting data on migration at a European level extend back to the mid-1970s, but it was only in the 1990s and in the context of ascendance of migration and asylum policy as a genuine EU policy field that a more systematic collection and dissemination of statistical data on migration was promoted. Thus, a programme for the collection of statistics on international migration was launched in 1992, and, in the mid-1990s linked to a joint data collection programme run in cooperation with the United Nations Economic Commission for Europe (UNECE), later also joined by the UN Statistical Division (UNSD), the Council of Europe (CoE) and the International Labour Office (ILO). In parallel, renewed efforts were undertaken to harmonise statistics on migration. Overall, data collection was undertaken largely on the basis of a “gentlemen’s agreement” lacking in timeliness, coverage and common definitions and a competence of the EU’s statistical agency to enforce standards. The adoption of the ‘Regulation on Community Statistics on Migration and International Protection’ (Regulation 862/2007) in July 2007 (since amended by Regulation (EU) 2020/851) was a major turning point in EU data collection on migration and asylum improving both availability and comparability of statistics and covering both administrative and demographic indicators. In regard to demographic statistics, the EU Regulation basically adopts the definitions of the 1998 UN recommendations and related census recommendations (see on concepts in more detail below).

27.2 Concepts and Definitions

In this section, we discuss concepts and definitions used for measuring migration and migrants and provide a heuristic classification of different conceptual approaches.

Migration can be broadly defined as *a specific form of spatial mobility, aimed at a certain minimum duration of residence in the destination and a similar minimum duration of residence in the place of origin*, thereby distinguishing it from other

Table 27.1 Concepts, indicators and categorisations used in migration-related statistics in European countries

Concept	Indicators/proxies	Examples of breakdowns/aggregations/ categorisations
Legal status	Citizenship	National/non-national
		EU-citizens/third country nationals
		Western/non-Western
	Residence permit/Work permit	Residence permit/no residence permit (irregular residence)
		Duration of validity
Purpose of residence permit (e.g. employment, education, family, international protection, . . .)		
Geographical origin	Country of birth/country of previous or next residence	Native/foreign
		EU/non-EU
		Western/non-Western
		Duration of residence in country
	Parents' country of birth	National/non-national
		EU-citizens/third country nationals
		Western / non-Western
Identity, culture, vulnerabilities	Parents' country of birth	National/non-national
		EU-citizens/third country nationals
		Western/non-Western
	Ethnicity/race	Autochthone/allochthone
		White/Black/Asian/ . . .
	Language (first language or language spoken at home)	Native/non-native
Religion	Muslim/Christian/ . . .	

Source: authors' elaboration based on Kraler et al. (2015), p. 51

forms of spatial mobility such as commuting or tourism. Political geography, reasons of migration, **forms**, and broader **drivers** of migration can be used as further differentiations.

According to the most recent UN definition of migration, migrants are defined as individuals who change their place of usual residence across international borders (United Nations, 1998) for at least 3 months. The definitions further distinguish between long-term migrants of an actual or intended duration of residence above 12 months and short-term migrants, with an intended or actual residence of between three and 12 months. Individuals changing their residence for less than 3 months are not counted as migrants.

There are many different ways in which migrants and related concepts are measured and counted across countries and even within countries. These practices are influenced by different historical experiences and political contexts (Fassmann et al., 2009). Table 27.1 provides an overview of different categorisations potentially used to measure migrants, migration and related concepts. These can be summarised very broadly, and not exclusively, as (1) individuals' legal status, including

citizenship or residence status, (2) individuals' geographical origin, including the country of birth or parents' country of birth, as well as (3) individuals' identity and/or culture, including ethnicity or language.

The concepts linked to identity and culture are challenging for at least two reasons. First, they might reflect racialised categories and thus unwittingly reinforce racism and discrimination. Secondly, they foreground specific categories and suggest that these are important in explaining particular outcomes, for example when presenting employment outcomes for persons with a "migration background" (see Simon et al. (2015) on controversies regarding such concepts). Following the recommendations of the United Nations Office of the High Commissioner of Human Rights, statistical measurements of ethnicity should be based on self-identification. In addition, the production of data about ethnic groups should use a participatory approach through involving affected groups in the process of producing such data (Office of the High Commissioner Human Rights, 2018). In the past decades, several countries moved away from measuring the ethnicity of people, often not based on self-identification, to more objective concepts such as country of birth. However, other countries introduced the collection of data based on race or ethnicity for the very purpose of measuring inequality and discrimination (e.g. Ireland or the United Kingdom). While there are concerns about the racialisation of society through the collection and dissemination of statistics on ethnicity, these statistics can serve the purpose to detect vulnerabilities and highlight the situation of vulnerable groups, which would otherwise be invisible to policymakers. The collection of ethnic statistics can be advantageous for the production of "equality data", which are needed for policies to detect and counter-act, discrimination and inequality across protected and vulnerable groups (European Commission, 2018).

The opportunities to collect statistics on the concepts described above, depends on the availability of the indicators and breakdowns used to measure those concepts. Various data sources can be used, which all have their advantages and disadvantages, as outlined in the next section.

27.3 Measurement and Data Sources

How can we measure migration? Basically, there are two different ways to look at migration and migrants: stocks and flows. Migration stocks refers to the number of migrants in a defined place at a given time point. Flows refer to events (i.e. migration movements) that take place within a given period of time. Stocks of migrants change over time due to in- and out-flows of migrants. Generally, stocks can be measured much more easily than flows, because migration dynamics are more difficult to capture.

A classic, vivid example of challenges linked to measuring migration flows is comparing so-called 'mirror statistics'. They compare the outflows from one country [COUNTRY A] to another country [COUNTRY B] with the inflows from that other

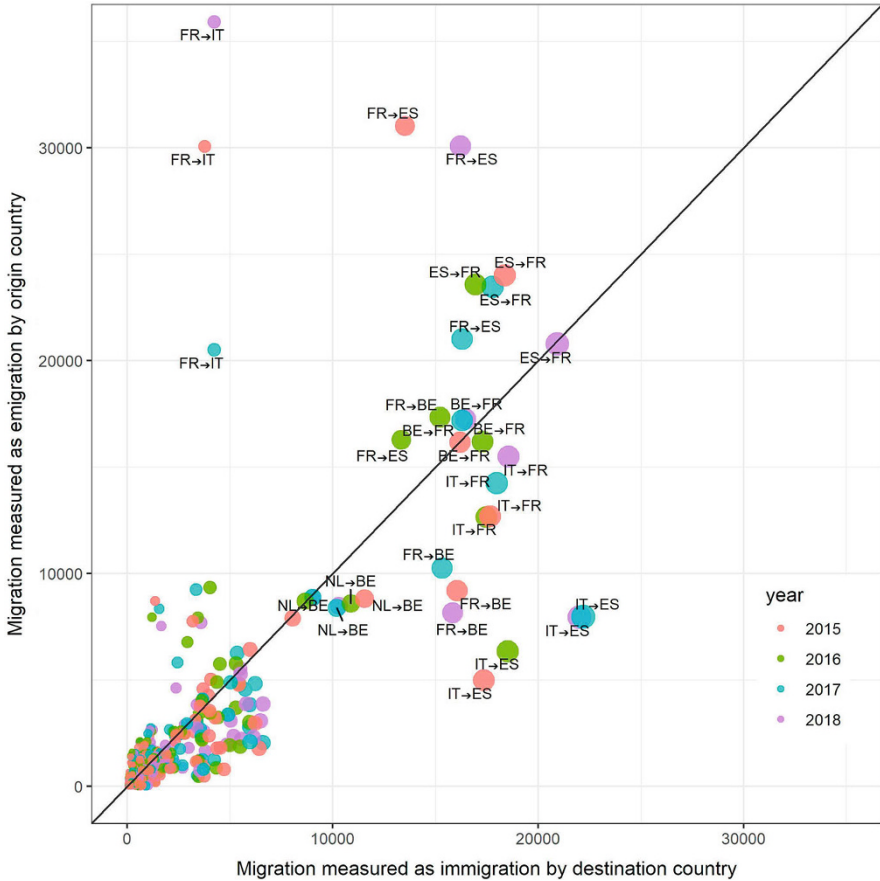


Fig. 27.1 Migration flows, as measured by origin and destination countries in selected European countries, 2014–2018. (Source: Authors’ analysis and presentation based on data from Eurostat, tables migr_imm5prv and migr_emi3nxt, data extracted on 24 October 2020)

country [COUNTRY B] from the one country [COUNTRY A]. Migration flows are usually measured based on citizenship or country of birth of those migrating. This information is only a proxy for the actual migration movement and does not directly indicate if a person moved from COUNTRY A to COUNTRY B. This means it can be assumed that most people born in Belgium moving to France are actual movements from Belgium to France. However, secondary movements are not captured this way. The European statistical office (Eurostat) provides two tables, which should capture direct migration flows for each EU Member State by country of previous and country of next residence.

Figure 27.1 shows migration flows from selected European countries and compares the numbers from 2014 to 2018 as measured by the emigration country (i.e. emigrants by country of next residence) against the numbers measured by the immigration country (i.e. immigrants by country of previous residence). If emigration and immigration were measured perfectly on both sides, sending and receiving country, the points in Fig. 27.1 would all be on the diagonal line from the bottom left to the top right. Between 2014 and 2018, Eurostat provides complete migration flows statistics including sending and receiving data for 15 sending and 15 receiving countries and altogether 200 flows, measured by both sides (400 numbers). The average emigration for each bilateral flow (averaged over the years 2014–2018) range from zero (SK to EE) to 24,560 (FR to ES) with an average number of about 1473. The average immigration numbers (i.e. which should be the same as the emigration numbers, but measured from the receiving country) range from three (EE to SK) to about 20,000 (IT to ES) with an average number of 1487. As emigration is more difficult to measure, usually immigration numbers are more reliable. For this set of bilateral flows, the immigration numbers are on average only slightly higher. There is an average difference of 14 and the number correlate at about 0.8. However, there are large differences observable in both directions. The largest average differences can be observed between France and Italy. Italy sees an average of 4058 immigrants from France between 2014 and 2018, but France counts on average 23,968 emigrants to Italy. This is an average difference of almost 20,000 people. On the other side, in terms of under-estimating emigration, which is more common, the most extreme case is the measurement of flows from Italy to Spain. Italy sees an average of 6818 people moving to Spain, but Spain registers an average of 20,000 people coming from Italy. Generally, there is good consistency in these numbers in these 15 countries. Since 2014 the numbers correlate at 0.78. However, countries with less reliable data most likely did not provide them to Eurostat if they did not meet the quality criteria.

Why is this so different? Much of the differences in migration flow statistics—but not everything—are linked to the ways data are collected. The main two sources for migration statistics are questionnaire-based data collections for the purposes of producing statistics, such as surveys, and administrative data sources, such as population registers.

Generally, the majority of EU countries use registers as the basis for producing migration flow statistics. Looking at all EU Member States, EFTA countries and the United Kingdom, only five use surveys (CY, IE, FR, PT and UK). Romania and Greece use other data sources including mirror statistics from other countries. The remaining countries use population registers, partly including registers for foreign citizens (e.g. DE).³ The remainder of this section discusses these data sources.

³See Eurostat metadata, Immigration (migr_imm), available at: https://ec.europa.eu/eurostat/cache/metadata/en/migr_imm_esms.htm, accessed 1 November 2020.

27.3.1 Questionnaire-Based Data Collection: Sample Survey and Traditional Census

One of the most important sources for population and social statistics—not only on migration—are questionnaire-based surveys (see also Salamońska (Chap. 26), this volume). Either a part of the population (sample surveys) or the entire population of interest receives a questionnaire including questions linked to their migration experiences and intentions. Usually, sample surveys are conducted because it is simply too expensive—and burdensome—for respondents to ask everyone of concern. To be able to make statements about the total population of interest, a random sample needs to be drawn, meaning a statistically controlled way of selecting respondents. This way the sampling error can be calculated, which provides an indication of the level of uncertainty in the statistics produced based on the samples.

However, statistical variation, due to the random selection from a list of all potential respondents, is not the only source of error in the production of statistics from sample surveys, as there are also issues linked non-response, question wording and categorisations (see Groves et al., 2009). Despite the many challenges, sample surveys have proved to be a reliable and much-used form of data collection for social statistics and research. The two largest EU-wide sample surveys are used for measuring employment and unemployment, the EU Labour Force Survey (EU LFS), and for income and living conditions, the EU Survey on Income and Living Conditions (EU SILC). However, the use of general population sample surveys remains challenging when it comes to measuring migration flows and stocks. This is mainly because migrants usually only make up a small share of the total population and are only captured in a limited way. Consequently, the sample sizes of migrants within general population surveys are often too low for statistically meaningful analysis. In addition, the sample of migrants might be biased, whereby certain groups of migrants cannot be captured well with methodologies used for general population surveys.⁴

To mitigate the lack of data about migrants from general population surveys, a variety of targeted surveys among migrant populations were carried out. This is sometimes complicated because in many countries, no sampling frames for migrants exist. This means that there are no lists of migrants, which are needed for a controlled way of selecting a representative sample. As a result, oftentimes alternative sampling methodologies need to be employed (see Reichel & Morales, 2017).

Traditionally, countries conduct a population census every 10 years. In a ‘traditional population census’ all with their usual residence in a country receive a questionnaire including questions about them. The majority of censuses in Europe contain questions on the country of birth and **citizenship** of people, which can good proxies for migrant stocks.

⁴For an analysis of availability of data for measuring migrant integration indicators see Huddleston et al. (2013).

Censuses are reliable sources for measuring migration stocks. However, a census can also leave out important groups of migrants due to difficulties in capturing migrants and varying or unclear definitions, for example [asylum seekers](#). The main disadvantage of traditional censuses is the infrequent availability of data because censuses are usually carried out every 10 years—due to their high cost. Costs are the main reason for many countries moving towards register-based censuses. This means that the information collected every 10 years about the population in a country is not obtained from questionnaires, but from information available in population and other administrative registers, sometimes used in combination with (linked) sample surveys. The use of registers is discussed in the next section.

27.3.2 *Administrative Data Sources*

Many statistics are produced based on administrative data, although the production of statistics is not necessarily the original purpose of data collection. For measuring migration, population registers are the most important source of statistics. The use of population registers is necessary to determine the place of residence of an individual, which is linked to several administrative and bureaucratic procedures, for example, which municipality is responsible for administering social welfare applications or voter registrations. However, the production of statistics is also one of the purposes of keeping population registers and hence some of the information collected and stored might only serve the purpose of producing statistics and not the administration of individuals. Thus, population figures are often needed for the allocation of resources within polities. Population statistics or specific subpopulations (such as third-country nationals) also are used for the distribution of financial resources in the EU, sometimes in combination with other indicators.

Apart from population registers, countries use data from registers containing information on residence permits and sometimes other registers, such as health or social security registers. Population registers are often maintained at municipal level.

Once in place, it is much cheaper for public administration to produce statistics based on administrative registers, as compared to separate questionnaire-based data collection. Challenges linked to the use of registers for migration statistics are:

- the lack of comparability across countries due to differences in administrative procedures, which do not necessarily follow any harmonised definition on migration, and
- the potential systematic undercounting of certain groups not captured in administrative registers (e.g. undocumented or unregistered migrants, groups not captured by certain administrative procedures).

27.3.3 *Big Data*

At the 2008 annual conference of the IMISCOE research network in Bilbao (Spain), demographer Michel Poulain gave a keynote speech on the general challenges of measuring migration stocks and flows. Asked about future developments in regard to measuring migration, he rather jokingly remarked that in the near future everybody will have a chip implanted that would allow following the exact movement of people in real time. Now, slightly more than 10 years later, this is almost true. Not that people have chips implanted, but the vast majority of people (in developed countries) carry a device with them, almost everywhere, that tracks their movement—a smart phone. What is more, people use new ways of communication over the internet and through social media, which also provides a source of information about migration movements (see also [ICT facilitating travel & migration](#)). Such data sources are often referred to as ‘big data’—data which are produced over the internet, including via smart phones and social media. Big data are usually characterised by increased velocity, volume, and variety compared to ‘traditional data’. This means data are quicker available, there are more observations (e.g. compared to sample surveys) and more information contained from a variety of different sources. This has the potential to add to traditional data collection and potentially mitigate challenges of measuring migration, as discussed above.

In the aftermaths of the large migration flows to Europe during 2015 and 2016 due to the war in Syria, efforts to better manage migration led to new initiatives to collect, compile, analyse and disseminate migration data. Most notably the International Organization for Migration (IOM) created the Global Migration Data Analysis Centre (GMDAC).⁵ At the level of the EU, the European Commission set up the Knowledge Centre on Migration and Demography (KCMD) within the Commission’s Joint Research Centre.⁶ Apart from collecting, analysing and disseminating a variety of migration related data and statistics, the two centres initiated the Big Data for Migration Alliance (BD4M). This initiative brings together different stakeholders to investigate “*the potential of big data sources for migration analysis and policymaking, while addressing issues of confidentiality, security and ethical use of data*” (GMDAC & KCMD, 2018). The Migration Data Portal includes a good overview of potential use of and initiatives related to big data for migration statistics, including potential sources of data and related challenges. Data sources for migration measurement from new data sources include data from mobile phones, the internet, and sensor (e.g. satellite imagery).⁷ Global Positioning System (GPS) tracking can be used via smart phones, as well as other location information provided by individuals for example via social media.

⁵<https://migrationdataportal.org/about>.

⁶https://ec.europa.eu/knowledge4policy/migration-demography_en.

⁷Migration Data Portal (2020), Big Data, migration and human mobility, last update 10 November 2020, available at: <https://migrationdataportal.org/themen/big-data-migration-and-human-mobility>.

One study looked into the use of Facebook data to measure international migration. This study used information about Facebook users' profile, provided by Facebook's advertising platform, concerning the countries a user used to live in. The study found that for selected groups of migrants in some countries, migration trends could be successfully and relatively accurately captured (Spyratos et al., 2019). Another study used data from the short messaging service Twitter, by looking into changes in the location information provided by its users' messages ('Tweets'). This is particularly challenging because users do not always provide information about their location on Twitter and the group of users is, as with other social media users, not representative of the total population. The study could still extract some information about migration trends (Zagheni et al., 2014).

These are just two examples. Several other studies have looked into the potential use of new data sources for migration measurement. For example, using big data to complement and improve mobility within the EU (Gendronneau et al., 2019), using Google search data for migration flows (Wanner, 2020b) or forced migration (Connor, 2017). Most recently, after the outbreak of the COVID-19 pandemic, data on mobility patterns were published by large companies running smart phone operating systems, Apple⁸ and Google,⁹ and also some cases for internal migration. While these data do not include information on origin and destination of movements, it shows people's behaviour in terms of general movement and mobility, and gives a sense about how fine-grained some data collections from these large companies are.

While many of the studies show promising result in terms of a potential added value using new data sources, such data collections are still in experimental phases or provide insight only for selected cases. The advantages, as mentioned above, are obviously additional information about migration flows, which cannot be gathered from traditional data sources, the ability to collect information much faster, particularly in comparison with questionnaire-based data collection, and the potential to better measure internal migration patterns at the same time as international migration. Yet, there are also challenging drawbacks. These are mainly the biased nature of big data and legal and ethical problems linked to privacy and data protection.

Bias

Bias means that the data do not measure exactly what they should measure—the data are not representative of the target population and consistently miss their target in specific direction. While most data collections are biased to some degree, big data sources are particularly prone to bias, because particular population groups are over-represented among users of different online services. Most notably, not all people in the world have access to the internet and even fewer own smart phones. According to the International Telecommunication Union, the percentage of individuals in a

⁸<https://covid19.apple.com/mobility>.

⁹[https://www.google.com/covid19/mobility/..](https://www.google.com/covid19/mobility/)

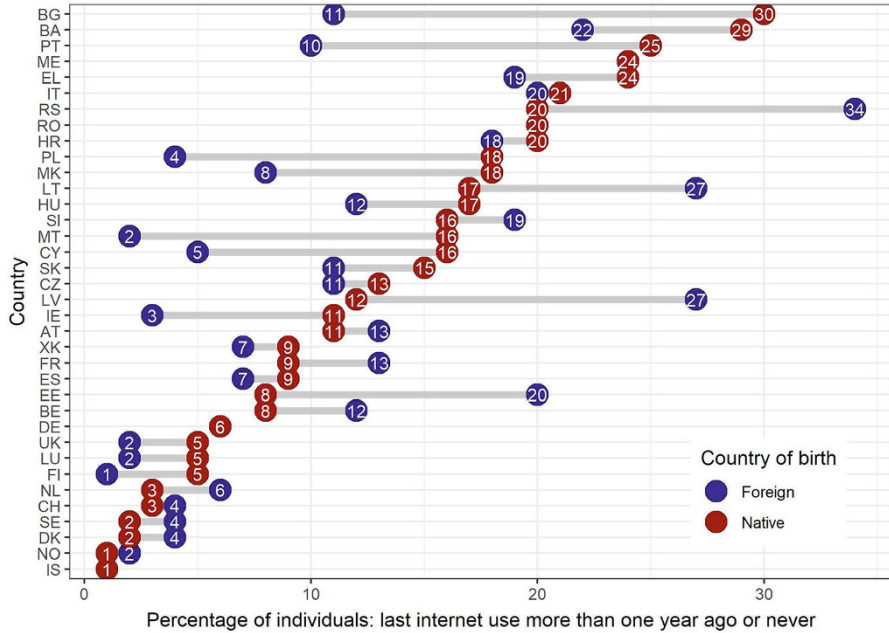


Fig. 27.2 Individuals not using the internet in the past year or never using the internet in European countries, by country of birth (in %). (Source: authors’ analysis and presentation based on data from Eurostat, table isoc_ci_ifp_iu, data extracted on 16 November 2020)

country who used the internet in the last 3 months ranges from 1.3% (Eritrea, in 2017) up to 99.7% (in Bahrain, in 2019). The median rate of internet access across the globe (using the most recent data from 2017 to 2019) was 60%.¹⁰ For many European countries data on internet use by country of birth area available. The data show that in some countries a considerably higher percentage among migrants does not have access to the internet compared to the native-born population. However, in other countries the share among migrants not using the internet is much lower compared to native born (see Fig. 27.2 for an overview of those without internet access by country of birth). This means that there is a different selection into access to the internet and hence different levels of representativity of data drawn from the internet. On top of that, the share of users of different web services and social media platforms, and the share of those also sharing certain data (e.g. location data) cannot be assumed to be equal across migrant groups. It is this bias that needs to be analysed and taken into account when considering the use of big data for producing migration statistics.

¹⁰Own calculations based on data from the World Development Indicators database, <https://data.worldbank.org/indicator/IT.NET.USER.ZS>. See computer code supplemented to the chapter, lines 180–209.

27.3.4 Privacy and Data Protection

Statistics drawn from big data sources are often based on personal data. Data are personal if they relate to an identified or identifiable individual person. People might think about data as anonymised if basic identifiers, such as names, social security identifiers, addresses, date of birth, are removed. However, studies have shown that with a certain amount of data relating to individuals, they can still be identified (Rocher et al., 2019). European Union (EU) and Council of Europe laws provide a strict framework for processing personal data. Accordingly, data are personal if they include information that can lead to re-identification considering someone applies reasonable means that are likely to be used for de-anonymisation (or re-identification) of the data. Thus, re-identification needs to be reasonably likely and not just theoretically possible. In the EU, whenever data are not anonymised, and hence contain personal data, the processing needs to follow the principles laid out in the General Data Protection Regulation.¹¹ These principles include lawfulness, fairness and transparency of processing, purpose limitation, data minimisation, data accuracy, storage limitation, integrity and confidentiality. In order to be lawful, personal data may only be processed if it is based on consent of the data subject, or required by law, for vital interests, public interest or some other legitimate interest. Importantly, sensitive personal data are subject to stricter rules. Sensitive data include data revealing the racial or ethnic origin of people and are hence more relevant for migration statistics. Importantly, if collected legally, further processing of data for statistical purposes is generally accepted as a compatible purpose, if the processing puts in place measures to secure the anonymisation (for a full overview of European data protection laws, see European Union Agency for Fundamental Rights (FRA) et al., 2018).

Big data sources include a variety of information, often linked to protected attributes and sensitive data, such as ethnic origin, political opinion and sexual orientation. Due care needs to be taken when using these data for the protection of privacy and personal data of individuals.

27.3.5 Machine Learning and Artificial Intelligence

In addition to the general production of statistics, the increased availability of data in combination with increased computing power led to considerable progress in the area of machine learning. Machine learning is a branch of artificial intelligence and often simply referred to as artificial intelligence. Machine learning is mainly used for

¹¹ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), <http://data.europa.eu/eli/reg/2016/679/oj>.

automating tasks based on statistical prediction rules. This use, referred to as supervised machine learning, works in a way that an existing data set that includes the desired outcomes is used to build prediction rules. A classic example is a spam filter. A data set with emails marked as spam or not spam ('ham'), where the machine learning algorithms searches for correlations among words and word combinations, as well as potentially other metadata attached to the emails, which are relevant for predicting if an email is spam.

There is considerable interest in using such technologies in the area of migration management. While the use of machine learning algorithms could also contribute to an improved production of migration statistics, it is currently tested to use migration statistics to predict migration inflows to the EU. For example, the European Asylum Support Office (EASO) works on ways to predict asylum inflows through big data for an early warning system (European Asylum Support Office, 2019, p. 51). Automation of process through machine learning algorithms can be beneficial to certain processes. However, it may have considerable impact on fundamental rights in various ways (European Union Agency for Fundamental Rights, 2020).

27.4 Analysing Aggregate Statistics on Migrants, Refugees and Non-nationals—Some Examples

In this final section of the chapter, we briefly demonstrate the use of aggregate migration statistics by way of example. We first present a brief analysis of global migrant stock statistics from the United Nations. Second, we combine it with data on the Gross Domestic Product (GDP) in the countries. Finally, we provide a short overview of refugee stocks data across the world and analyse of the share of refugees among migrants.

The analysis can be reproduced by using the free statistics software R (R Core Team, 2019) and all the code for the analysis is available.¹² The code can be used and adapted or further developed for any further analysis.

27.4.1 Global Migrant Stocks: Main Countries of Origin and Destination

In 2019, the United Nations estimated the number of migrants at around 272 million. By far, most migrants are living in the United States of America, with 50 million or 19% of all migrants. Other major destination countries include Germany and Saudi Arabia, each with around 13 million migrants (or 5% of all migrants), Russia, the

¹²<https://github.com/david-reichel/migration-statistics>.

United Kingdom, United Arab Emirates, France, Canada, Australia and Italy (in descending order from 12 million to about 6 million migrants).

The most important countries of origin are India, with 17.5 million migrants or 6% of all migrants, Mexico (11.8 million and 4%), China, Russia, and Syria (8 to 11 million). When analysing the global migrant stocks over time, Czaika and de Haas (2014) find that the world has not necessarily become more migratory over the past decades (1960–2000), but more globalised in terms of destination countries, yet not from origin country point of view.

This is also the case when looking into migration stocks for 2019. While the ten most important destination countries host already over 50% of all migrants, the ten most important countries of origin account for slightly more than one third of all migrants (34.2%). It is particularly richer ‘Western’ countries that show more diverse countries of origin, whereas the main destination countries in the global South usually only count a few countries of origin. This is exemplified in Fig. 27.3, where the (logged) number of migrants in all countries is plotted against the number of origin countries in the destination countries. The names of those with more than 2.5 million migrants are highlighted.

The total number of migrants in a country is obviously linked to the size of a country with respect to the total population, because larger country might more easily host more migrants. Overall, there were about 272 million international migrants in 2019, which is a proportion of 3.5%. This indicates, in fact, that migration is rather the exception, than the rule, with less than three in 100 people living in another country. In some countries, the share of migrants is miniscule compared to the total population constituting less than 1%. However, this is not necessarily a small number of migrants, for example China and India host an estimated number of one and five million international migrants, respectively, but these still makes up less than 1% of the total population. The countries with the largest share of international migrants in 2019 was the United Arab Emirates where the 8.6 million migrants constitute 88% of the total population.

27.4.2 Immigrant and Emigrant Rates by Productivity

On a macro-level, the level of economic productivity of a country has long been linked to migration rates (see Czaika & Reinprecht (Chap. 3), this volume). The higher a country’s productivity the higher the rate of immigrants among its total population. Interestingly, this does not apply vice versa to the rate of emigrants. Here the emigration rate is very low among low productivity (i.e. poorer) countries, is higher among medium productivity countries and then decreases again. This phenomenon is sometimes referred to as the ‘migration hump’ because of the shape of the statistical development of the average emigrant rate by GDP (see Fig. 27.4).

Importantly, this pattern describes a cross-sectional phenomenon and must not be mistaken with a causal relationship. This means that if a country moves from a low to a medium GDP country, the number of emigrants does not necessarily increase, and

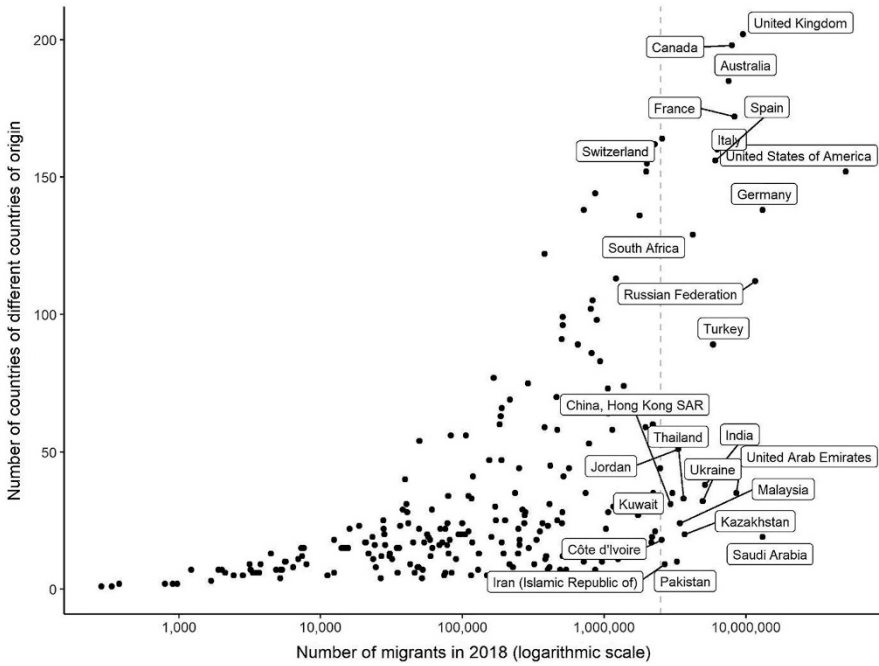


Fig. 27.3 Number of migrants and number of countries of origin in destination countries, 2018. (Source: own calculation and presentation based on data from United Nations 2020. Countries with more than 2.5 million migrants are labelled. Number of origin countries are for countries with more than 100)

the other way round. It is, however, a good reminder that the economy is an important driver of migration, and that people in the poorest countries not necessarily send many migrants around the globe (Benček & Schneiderheinze, 2019).

The Role of Forced Displacement in International Migration

Overall, there were around 3.5% migrants across the globe in 2019. However, the percentage of migrants varies considerably across countries, ranging from close to 0% up to about 88% in the United Arab Emirates. While discussions about migration often focus on issues linked to refugees, at around 20.3 million refugees in 2019 they only make up a relatively small share of the global population (just under 0.3%) and some 7.5% of the total global migrant population.¹³ However, refugee stocks are

¹³ It is important to note, and as discussed in section, that refugee data are separately collected from migrant data, and hence, there can be some differences.

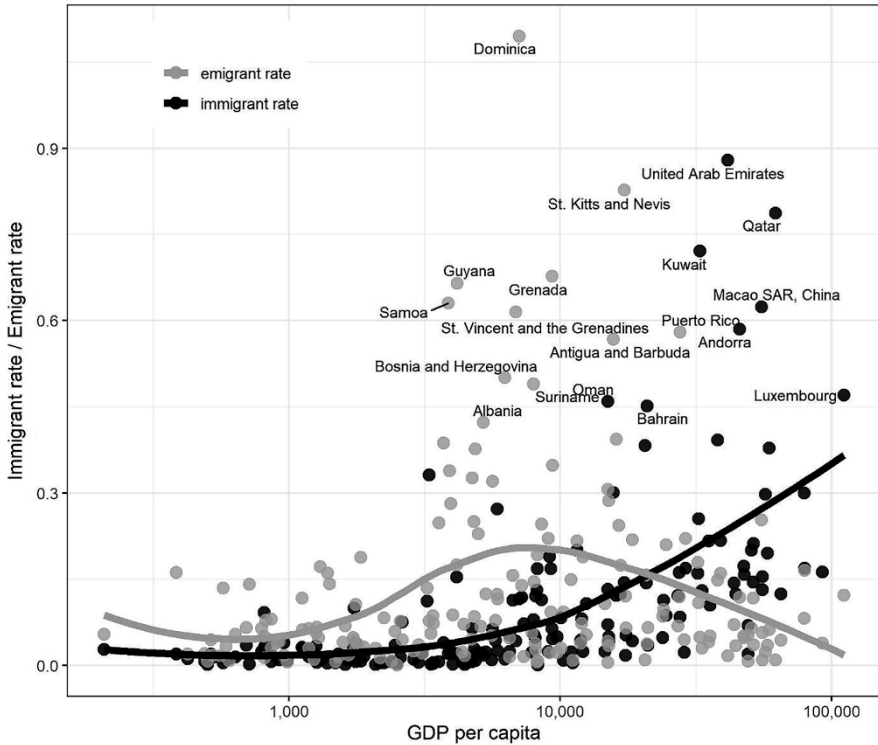


Fig. 27.4 Immigrant and emigrant rates by GDP, 2019. (Source: own calculation and presentation based on data United Nations, 2020 and World Bank Indicators (NY.GDP.PCAP.KD) via Vincent, 2021)

distributed highly unequally and in some countries—usually poorer countries—forced displacement dominates migration dynamics.

Globally, the percentage of refugees in the total population ranges from virtually zero up to 14% in Lebanon. In a similar vein, share of refugees in the total number of international migrants ranges from close to zero to up to 88% in Sudan and Chad. There are 12 countries where the share of refugees makes up at least half of the total migrant population of the country. The total number of refugees in these 12 countries, accounts for about half of all refugees worldwide. These are countries with considerably lower levels of production and income, as measured through GDP per capita and shown in Fig. 27.5.

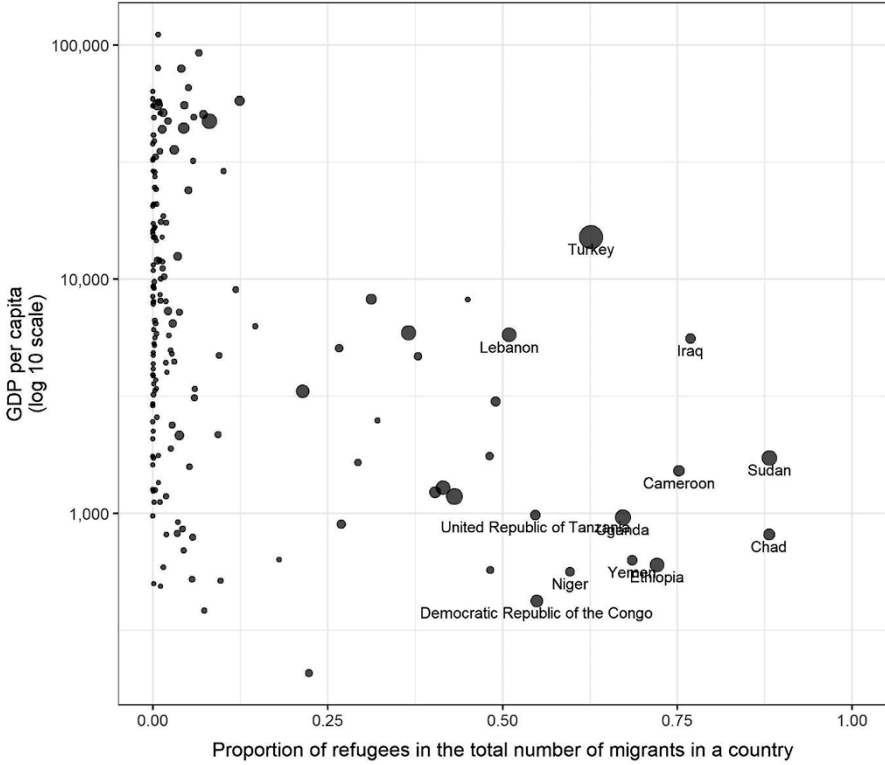


Fig. 27.5 Share of refugees among migrants and GDP, 2019. (Source: authors’ calculations based on data from UNHCR, 2020, United Nations, 2020 and World Bank Indicators (NY.GDP.PCAP.KD) via Arel-Bundock, 2021)

27.5 Conclusions

As this chapter has shown, (aggregate) migration statistics play an important role in academic and policy debates on migration and they remain a central tool for analysing, understanding, and debating migration and related issues. The increasing availability of a growing range of migration indicators that can be accessed through different databases and in different formats and disaggregations, provides ample opportunities for analysing migration dynamics and relate them to a wide range of other macro-level indicators (including policy indicators, see Solano and Huddleston (Chap. 24), this volume).

Importantly, aggregate indicators are never sufficient and meaningful in themselves, as they always require interpretation and are often of uncertain quality that has to be critically assessed. Refugee statistics that are briefly analysed above are a case on point, for which only recently global standards have been defined. But also,

in regard to core migration indicators for which standards have first been defined in the 1930s, many issues remain, especially on the global level (see Ahmad-Yar & Bircan, 2021). The user thus needs to be aware of both what and how data is collected, in order to understand what information they can provide. Despite deficiencies of migration statistics, they are however an indispensable prerequisite and indeed a powerful tool for broader macro-level generalisations¹⁴. Ernest Ravenstein's *Laws of Migration* (Ravenstein, 1885)—one of the foundational studies of Migration Studies—is a vivid example of the potential of the analysis of aggregate statistics—to derive hypotheses from empirical observations and prepare the ground for further empirical and theoretical studies (See also Carling et al., 2020).

While the ICT revolution means that each individual leaves millions of data traces in using modern communication devices, the potential of 'big data' and other new data sources for analysing macro-level migration dynamics remains unclear and subject to further research. While there are indeed a growing number of studies providing specific insights and demonstrating the potential of new data sources, a wider application of big data that would provide regular, comparable, and reliable indicators is currently still out of reach. In addition, there are also important privacy and data protection issues that need to be addressed. Yet, if handled properly, such data sources will become an increasingly useful source to complement and enrich traditional data collection on migration.

Bibliography

- Ahmad-Yar, A. W., & Bircan, T. (2021). Anatomy of a misfit: International migration statistics. *Sustainability*, 13, 4032. <https://doi.org/10.3390/su13074032>
- Andreas, P., & Greenhill, K. M. (2010). *Sex, drugs, and body counts: The politics of numbers in global crime and conflict*. Cornell University Press.
- Arel-Bundock, V. (2021). WDI: World development indicators and other World Bank data. R package version 2.7.2. <https://CRAN.R-project.org/package=WDI>
- Behrmann, E. (2013). Gates helps Australia's richest man in bid to end slavery. *Bloomberg*, 15.4.2013. <https://www.bloomberg.com/news/articles/2013-04-10/gates-helps-australia-s-richest-man-in-bid-to-end-slavery>
- Benček, D., & Schneiderheinze, C. (2019). More development, less emigration to OECD countries – Identifying inconsistencies between cross-sectional and time-series estimates of the migration hump. 2145.
- Bijl, R., & Verweij, A. (2012). *Measuring and monitoring immigrant's integration in Europe. Comparing integration policies and monitoring systems for the integration of immigrants and ethnic minorities*. Netherlands Institute for Social Research.
- Boswell, C. (2018). *Manufacturing political trust*. Cambridge University Press.

¹⁴Wanner (2020a) has first made this point in his recent analysis of statistical data collection on migration.

- Carling, J., Czaika, M., & Erdal, M. B. (2020). Translating migration theory into empirical propositions. *QuantMig Project Deliverable 1.2*. <http://www.quantmig.eu/res/files/QuantMig%20D1.2%20v1.1.pdf>
- Connor, P. (2017). Can Google trends forecast forced migration flows? Perhaps, but under certain conditions. (Paper presented at population Association of America meetings, April 2017). Chicago, Illinois. <https://paa.confex.com/paa/2017/meetingapp.cgi/Paper/9477>
- Crisp, J. (1999). "Who has counted the refugees?" UNHCR and the politics of Numbers. New Issues in Refugee Research, Working Paper 12. <https://www.refworld.org/docid/4ff58e4b2.html>
- Czaika, M., & de Haas, H. (2014). The globalization of migration: Has the world become more migratory? *International Migration Review*, 48(2), 283–323. <https://doi.org/10.1111/imre.12095>
- Czaika, M., & Kraler, A. (2020). Refugees. In J. Stone, R. Dennis, P. Rizova, & X. Hou (Eds.), *The Wiley Blackwell companion to race, ethnicity, and nationalism* (pp. 321–348). Wiley Blackwell.
- European Asylum Support Office. (2019). *Annual report on the situation of Asylum in The European Union 2018*.
- European Commission. (2018). Guidelines on improving the collection and use of equality data.
- European Union Agency for Fundamental Rights (FRA). (2020). *Getting the future right. Artificial intelligence and fundamental rights*. Publications Office of the European Union. <https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Technical-Report/national-reporting/Technical-report-on-statistics-of-IDPs-E.pdf>
- European Union Agency for Fundamental Rights (FRA), European Court of Human Rights Council of Europe, and European Data Protection Supervisor (EDPS). (2018). *Handbook on European data protection law* (2018 edition). Publications Office of the European Union.
- Eurostat. (2018). *Technical report on statistics of internally displaced persons: Current practices and recommendations for improvement*. Publications Office of the European Union.
- Eurostat/Expert Group on Refugee and Internally Displaced Persons. (2018). *International recommendations on refugee statistics*. Publications Office of the European Union. https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/International-Migration/2018_1746_EN_08-E.pdf
- Fassmann, H., Reeger, U., & Sievers, W. (2009). Statistics and reality: Concepts and measurements of migration in Europe. Amsterdam University Press.
- Gallagher, A. T. (2017). What's wrong with the global slavery index? *Anti-Trafficking Review*, 8(2017), 90–112.
- Gendronneau, C., Wisniowski, A., Yildiz, D., Zagheni, E., Florio, L., Hsiao, Y., Stepanek, M., Weber, I., Abel, G., & Hoorens, S. (2019). Measuring labour mobility and migration using big data.
- GMDAC & KCMD. (2018). Big data for migration Alliance (BD4M). Harnessing the potential of new data sources and innovative methodologies for migration. Available at: <https://knowledge4policy.ec.europa.eu/sites/default/files/big-datafor-migration-alliance-concept-note.pdf>. (16 January 2022).
- Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (Second ed.). Wiley.
- Heller, C., & Pécoud, A. (2020). Counting migrants' deaths at the border: From civil society counterstatistics to (inter)governmental recuperation. *American Behavioral Scientist*, 64(4), 480–500. <https://doi.org/10.1177/0002764219882996>
- Herm, A. (2006). Recommendations on international migration statistics and development of data collection at an international level. In M. Poulain, N. Perrin, & A. Singleton (Eds.), *THESIM. Towards Harmonised European Statistics on International Migration* (pp. 77–106). Presses Universitaires de Louvain.
- Huddleston, T., Niessen, J., & Tjaden, J. D. (2013). Using EU indicators of immigrant integration.
- Kraler, A., Reichel, D., & Entzinger, H. (2015). Migration statistics in Europe: A core component of governance and population research. In P. Scholten, H. Entzinger, R. Penninx, & S. Verbeek (Eds.), *Integrating immigrants in Europe. Research-policy dialogues* (pp. 39–58). Springer.

- Kraly, E. P., & Gnanasekaran, K. S. (1987). Efforts to improve international migration statistics. A historical perspective. *International Migration*, 21(4), 967–995.
- Office of the High Commissioner Human Rights. 2018. A human rights-based approach to data. R Core Team. (2019). R: A language and environment for statistical computing. R foundation for statistical computing.
- Ravenstein, E. G. (1885). The laws of migration. *Journal of the Royal Statistical Society*, 48, 167–277.
- Reichel, D., & Morales, L. (2017). Surveying immigrants without sampling frames - Evaluating the success of alternative field methods. *Comparative Migration Studies*, 5(2017), 1.
- Rocher, L., Hendrickx, J. M., & Alexandre de Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10(1).
- Scott, J. C. (1998). *Seeing like a State*. Yale University Press.
- Simon, P., Piché, V., & Gagnon, A. (2015). *Social statistics and ethnic diversity: Cross-national perspectives in classifications and identity politics* (IMISCOE Research Series). Springer.
- Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2019). Quantifying international human mobility patterns using Facebook network data. *PLoS One*, 14(10).
- Starr, P., & William, A. (1987). *The politics of numbers*. Russell Sage Foundation.
- Torpey, J. C. (2000). *The invention of the passport : Surveillance, citizenship, and the state*. Cambridge University Press.
- UN Statistical Division. (2018). Update of recommendations for international migration statistics (concept note). Paper presented at the UNECE-Eurostat Work Session on Migration Statistics, 24–26 October 2018 in Geneva. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.10/2018/mtg1/UNSD_Recommendations_ENG.pdf
- UNHCR. (2010). *2009 global trends refugees, asylum-seekers, returnees, internally displaced and stateless persons*. UNHCR. <https://www.unhcr.org/4c11f0be9.pdf>
- UNHCR. (2015). *Global trends. Forced displacement in 2014*. UNHCR. <https://www.unhcr.org/statistics/country/556725e69/unhcr-global-trends-2014.html>
- UNHCR. (2020). What is a refugee? Retrieved July 20, 2020, <https://www.unhcr.org/what-is-a-refugee.html>.
- United Nations. (1998). Recommendations on statistics of international migration. Revision 1.
- United Nations Department for Social Affairs. (1949). *Problems of migration statistics*. Population Studies No. 5.
- United Nations Population Division. (2020). International migrant stock 2019, available at: https://www.un.org/en/development/desa/population/migration/data/estimates2/data/UN_MigrantStockByOriginAndDestination_2019.xlsx. Accessed in Aug 2020.
- Wanner, P. (2020a). Collection and analysis of quantitative data in the field of migration. Past trends, current status and future prospects. NCCR on the Move, Working Paper no 27. https://nccr-onthemove.ch/wp_live14/wp-content/uploads/2020/10/nccrotm-WP27_Wanner.pdf
- Wanner, P. (2020b). How well can we estimate immigration trends using Google data? *Quality and Quantity*.
- Zagheni, E., Kiran, V. R., Weber, I., State, B. (2014). Inferring international and internal migration patterns from twitter data. In WWW '14 companion: Proceedings of the 23rd International Conference on World Wide Web: April 7–11, 2014, Seoul, Korea.

Albert Kraler is Assistant Professor of Migration Studies in the [Department for Migration and Globalization at Danube University Krems](#). A political scientist by background, his current research focuses on forced migration and global asylum, and migration governance. Other research interests include migration statistics, regularisation and family migration policy.

David Reichel is a researcher at the EU Agency for Fundamental Rights (FRA), where his work focuses on artificial intelligence, big data and fundamental rights. His areas of expertise include statistical data analysis, data quality and statistical data visualisation, migration and integration statistics, citizenship, and human rights.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

