Reliable, timely, comprehensive data are critical to effective policy making, particularly for complex and polarizing issues such as migration. The Global Compact for Safe, Orderly and Regular Migration explicitly states as its first objective collecting and utilizing “accurate and disaggregated data as a basis for evidence-based policies.” Yet the current migration data landscape is characterized by divergences and inconsistencies in definitions, data collection methodologies, and dissemination strategies. The available data have large gaps across countries and over time stemming from irregular collection efforts, lack of harmonization, or failures in dissemination. Many of the data sets, including administrative data collected by governments, are beyond the reach of academics and policy makers.

Why data?

Data are needed to inform policy making for all types of movements and at all stages. For example:

- Migration data help policy makers better understand the magnitude of migration, its drivers (including the potential role of climate change), as well as its impacts. High-quality data allow analyses of who moves, with what skills and attributes, and under what circumstances.
- In origin countries, policy makers are interested in measuring how remittances can contribute to poverty reduction—which migrants contribute the most and with what patterns; how remittance flows affect poverty among various households in receiving communities; and how specific policy measures can enhance their effects. Other policy makers are concerned about the downside of emigration: What happens to family members left behind, and how can the issues they are facing best be mitigated? What are the actual effects of a brain drain in a given situation, which professions are most affected, and what is the effect of specific mitigation measures? Yet others look at the impact of emigration—including diaspora and return—on a country’s development.
- In destination countries, some policy makers are concerned about economic impacts—migrants’ skills and attributes, their participation in the labor market, their effects on productivity, the consequences for various groups of national workers, and the impacts of inclusion policies. Others are interested in social impacts—migrants’ abilities to integrate and the pace at which they do so, the effects on the provision of public services, and the differentiated impacts of various policy approaches to manage this process. Information on the subnational distribution of migrants can be critical to informing such discussions.
- Additional data related to forced displacement, transit, undocumented, and “distressed” migration patterns are also important for the relevant countries.
Definitions

Consistency of definitions—across data sources within a country, across countries, and over time—is essential for the effective use of migration- and forced displacement–related data. However, there are wide variations, including in high-income countries with solid statistical systems. For example, the Norwegian authorities and the United Nations High Commissioner for Refugees (UNHCR) noted significant discrepancies between the numbers of refugees in Norway reported for the end of 2013—18,734 by Eurostat; 46,033 by UNHCR; and 132,203 by Statistics Norway—reflecting differences in definitions, time frames, and statistical methods. Similarly, the definition of migrant in national censuses varies across countries. It can be based on place of birth, citizenship, time of arrival, or even ethnicity or race (figure S2.1). Countries, government agencies within the same country, and researchers collect data using any one of these definitions, making comparisons and analyses difficult.

Figure S2.1 Many population censuses do not collect basic and consistent data on migration

![Figure S2.1](image-source)

Source: Juran and Snow 2018. Based on data for 149 countries from United Nations Department of Economic and Social Affairs (UN DESA).

Note: The number of censuses refers to censuses in the 2010 round whose questionnaires include two or three core questions on country of birth, citizenship, and year of arrival.

Data sources

Multiple sources of data can inform policy making, each with its strengths and limitations. Each set of instruments is insufficient to fully address the data challenges related to migration. However, each can help provide evidence on specific aspects of cross-border movement. Effective policy making requires using multiple data sources across geographic, disciplinary, and institutional boundaries. Various data sources can complement one another so that, over time, a comprehensive picture emerges:
• **Population censuses** remain the primary data sources on global migration. On the one hand, their extensive coverage and granularity enable measurement of long-term movements of migrants. Population censuses are also conducted in almost every country at all income levels, enabling some level of comparability across countries and time periods. On the other hand, there are challenges related to their timeliness and accessibility. Censuses are most often conducted once every 10 years, and so they cannot capture rapid changes in migration patterns. In destination countries, censuses tend to undercount certain subpopulations, such as undocumented migrants, and other hard-to-reach populations, such as refugees. Even when coverage is not a constraint, information on citizenship is rarely collected, thereby leaving the citizenship of foreign-born people unknown and making it impossible to construct statistics on noncitizens. In origin countries, censuses cannot capture emigrants when the entire household has left. The issues covered in questionnaires are necessarily constrained by issues related to costs and quality, even when extended questionnaires are used for subsample groups. Finally, in some countries, political considerations and financial constraints limit access to census data.

• **Population registers** are another primary source of data on global migration. Similar to censuses, they cover a large part of a country’s population and provide long-term data sets. However, they are largely limited to higher-income countries. Population registers often fail to fully account for undocumented migrants and other marginalized populations who have little incentive to register. They may overestimate the numbers of migrants and refugees if those who leave fail to de-register—for example, if registration is tied to receiving benefits. Population registers are often managed by line ministries, which are peripheral to the national statistical system, and this factor limits accessibility. Canada, Spain, and Nordic countries are in the early stages of connecting population registers with a range of administrative databases.

• **Administrative data** are among the most promising yet underutilized sources for migration research and policy making. On the one hand, almost every country collects administrative data, but through different systems. Data are collected on who crosses their borders, taxation, social and welfare programs, pensions, health care, education, and other public services, among other things. On the other hand, ministries and agencies have little incentive to share, harmonize, or integrate their data sources due to national security, privacy, or bureaucratic concerns. The Organisation for Economic Co-operation and Development (OECD) is leading efforts in its member countries and other parts of the world to address some of these issues. However, the true value of these data will be realized only when data sources are integrated, such as through national ID numbers, while ensuring adequate levels of data privacy.

• **Household and labor force surveys** capture rich information on different aspects of migration, such as individual drivers, socioeconomic characteristics, and labor market impacts. Surveys in destination countries provide evidence on the extent to which migrants themselves directly benefit from their movement. Surveys in origin countries help shed light on the development impacts of mobility, particularly on the families and communities of emigrants. However, these surveys are conducted infrequently, and they are rarely conducted among a sample large enough to capture meaningful information about migrants. Expanding coverage is very costly, even more so if coverage is being expanded so that data for the same individuals are collected over multiple periods. In the short term, adjusting existing general survey programs to include more migration-related questions may be a more cost-effective option to improve the availability of data. Experimental approaches (such as impact evaluations) could also solve many of the remaining challenges.
Surveys for impact evaluations can help assess the effects of specific policies on various groups—migrants, refugees, or nationals—and fine-tune policy design and implementation. On the one hand, they provide the more direct evidence that can be used for policy making. On the other, the corresponding data are often costly to collect—if they do not rely directly on administrative data—with possible coverage challenges.

New data sources have been heralded as a solution for the many limitations of traditional sources. Mobile phone call records, geotagged social media data, internet traffic, and Internet Protocol (IP) addresses are being used to track forced displacement, predict migration trends, and analyze remittances. However, these sources suffer from their experimental nature, low statistical rigor, and biases in sampling. Their access for academic and policy research purposes can also raise privacy-related issues.

Additional challenges

Data collection on refugees and other marginalized migrant populations faces particular challenges. The sudden nature of some of these movements and the fact that they are occurring in areas often difficult to access or where administrative capacity is low complicate efforts considerably. At times, security and political considerations hamper data collection efforts as well.

The exclusion of forcibly displaced people from national statistical systems further marginalizes their access to social safety nets, public services, and employment opportunities. Some countries such as Chad have incorporated these populations into existing national data collection efforts. Chad’s Refugees and Host Communities Household Survey is fully integrated into its national household surveys as an integral part of the refugee policy dialogues among the government, the West African Economic and Monetary Union (WAEMU), and development partners.

Global efforts to improve the availability and quality of data on refugees and displaced populations have led to some improvements over the last decade. The global survey program on refugees by UNHCR, as well as initiatives such as the World Bank–UNHCR Joint Data Center on Forced Displacement (JDC), have demonstrated the feasibility and benefits of integrated data collection efforts. The Expert Group on Refugee, IDP and Statelessness Statistics (EGRISS), established by the United Nations in 2016, has developed two sets of methodological recommendations—one on refugee statistics and one on internally displaced person (IDP) statistics—that countries should implement to strengthen data collection efforts at national levels.

Data privacy

In a world characterized by a growing need for more reliable statistics and comprehensive, timely data, as well as the emergence of new technologies, data protection is becoming increasingly relevant. The collection of data on migrants, who often lack the legal protections granted to citizens in their country of residence, raises many privacy concerns—and even more so when migrants have an undocumented status. These concerns need to be addressed. Collecting, sharing, and processing data—even if it can help inform policy making—pose privacy risks, including data theft, data loss, and unauthorized use of personal data. In some sensitive contexts such as refugee situations, child migration, human trafficking, and migrant smuggling, confidentiality of personal data is especially important because identification of a data subject can pose life-threatening risks.
Governments and other public authorities should aim to address these concerns by using the best resources available when collecting migration- and forced displacement–related data, while ensuring the fundamental privacy rights and safety of data subjects. They should also ensure that applicable data protection and privacy laws adequately cover noncitizen populations; that they are effectively enforced, including for noncitizen populations; that safe data handling procedures are developed to prevent third parties from accessing potentially harmful data, including through anonymization procedures; and that migrants know their rights and are informed about the purpose of data collection and about data-sharing risks and mitigation measures. Similarly, when using new technologies, governments should ensure that they are transparent about how the data are being used to make decisions about migrants.

Looking ahead

Good data on migration are essential for governments to better manage migration and inform sound and effective policies. Yet there are significant data gaps, and major efforts are needed to collect both timely and granular data in a consistent manner. This will require more financial resources—including funding for long-term data collection—and support to strengthen the technical capacity of each country’s statistical agency—including lower-income countries. Dedicated data financing instruments such as the World Bank’s Global Data Facility can help address existing funding gaps while promoting greater coherence in investments in migration data.

Within this context, priorities to enhance the availability and quality of data that can inform policy making include:

- **Harmonization.** Efforts are needed to improve the consistency of definitions, or at least to collect data that can make comparisons possible even if countries use distinct definitions. For example, censuses should include at least four core questions: on country of birth, country of citizenship, country of previous residence, and year of arrival. Harmonization efforts undertaken by the United Nations Statistical Commission and OECD are critical. In addition to governments, many international and regional organizations collect data, particularly in areas where government data may be lacking such as on refugees. Coordination among the various actors is needed to prevent duplication, actively seek synergies, and ensure consistency in the methodologies used across surveys, thereby enabling comparability.

- **Innovative surveys.** Beyond traditional data sources, additional surveys can be conducted to inform policy making, including impact evaluations and surveys on the drivers and impacts of different types of movements. For example, longitudinal studies track migrants across borders and over time to understand the impact of migration or integration processes over time. Notable examples include the Mexican Family Life Survey and the Long-Term Impacts of Migration Survey on Tongan Migrants in New Zealand. Agreements across institutions in origin, destination, and transit countries are sometimes needed, such as for the Mexican Migration Project. Similarly, survey instruments are needed to analyze rapid and short-term movements using existing statistical systems, including movements lasting for short periods of time such as those in transit countries.

- **Data accessibility and privacy.** Facilitating access to data while ensuring confidentiality often requires a combination of actions, including strengthening legal and regulatory instruments that regulate data exchanges; enhancing administrative data systems; establishing legal rules, data license agreements, and shared secure architecture to facilitate the exchange of privately owned data; and developing standard data access agreements between the owners and the users of the data, such as the members of the research and policy-making community.
Notes

2. World Bank (2017). The UNHCR estimate is based on the total number of asylum-seekers granted a positive decision on their asylum claim over the last 10 years. The Eurostat estimate is based on valid residence permits issued to those granted refugee status or subsidiary protection. Finally, the Statistics Norway estimate is based on the number of “principal applicants”—179,534, when including persons who have been given a residence permit because of a family relationship with refugees.
5. Indeed, assessing the overall extent of undercounting through demographic analyses and postenumeration surveys is key to determining the credibility and usefulness of population censuses for measuring migrant stocks (Kennel and Jensen 2022).
6. See Poulain and Herm (2013) for a list of population registers available.
11. Bilsborrow (2016); Bilsborrow et al. (1997); Borjas (1987); de Brauw and Carletto (2012); Eckman and Himelein (2022); Heckathorn (2002); Kish (1965); McKenzie (2012); McKenzie and Mistaen (2007); McKenzie, Stillman, and Gibson (2010); McKenzie and Yang (2010).
12. Aiken et al. (2021); Hughes et al. (2016); Kim et al. (2020); Laczko and Rango (2014); Sirbu et al. (2021); Tjaden (2021); UNHCR (2021).
27. Rubalcava et al. (2008).

References


