The Future of Population Data amidst Demographic Megatrends

Alessio Cangiano

Senior Adviser on Demographic Megatrends

UNFPA Population and Development Branch, New York

Rationale and approach

In 1994, the International Conference on Population and Development (ICPD) recognized the fundamental importance of population and development data for government planning and monitoring, putting special emphasis on the need for more accurate data to define women's social and economic status and the needs of the most marginalized subpopulations. Since then, recurring calls for improving data systems and capacities have echoed these recommendations, including the United Nations Secretary-General's call for a "data revolution" to track and achieve the global Sustainable Development Goals (SDGs).

Based on contributions elicited from selected international experts[1] and information gathered through UNFPA's country and regional level engagement with National Statistical Offices and expert consultations under the aegis of the United Nations, this paper takes stock of the progress and extant gaps in population and development data and offers a critical outlook on the future of population data ecosystems in low- and middle-income countries (LMICs). It does so against the backdrop of demographic changes, other global megatrends and key political-economic shifts that will continue to transform the data value chain. We propose pathways towards better integrated population data systems, grounded in stronger registry-based solutions, and capitalizing on new geospatial and non-traditional data sources. We call for urgent attention to data governance frameworks, and expanded investments to meet the information and data protection needs for the post-2030 development agenda.

Progress and gaps in population and development data

Field-based population and housing censuses remain the cornerstones of demographic data in most LMICs, generating population denominators for more than 100 indicators to track the SDGs. Censuses are the primary (and often only) source of disaggregated data on small areas and marginalized subpopulations. SDG monitoring also relies heavily on comparable data generated by global household survey programmes that have grown in popularity and scope in recent decades.

The advent of multiple technological innovations has modernized data collection, production and use. In particular, the adoption of geospatial solutions across the census cycle has propelled data capture and analysis at highly granular levels and allowed data dissemination through maps that visualize

[1] The author gratefully acknowledges the contributions by Jacob Adetunji (Gates Foundation), Lorant Czaran (United Nations Office for Outer Space Affairs), Ridhi Kashyap (University of Oxford), Steve MacFeely (WHO), Romesh Silva (UNFPA), Rachel Snow (UNFPA Consultant), and Seiffe Tadesse (UN Statistics Division) that informed the preparation of this paper.

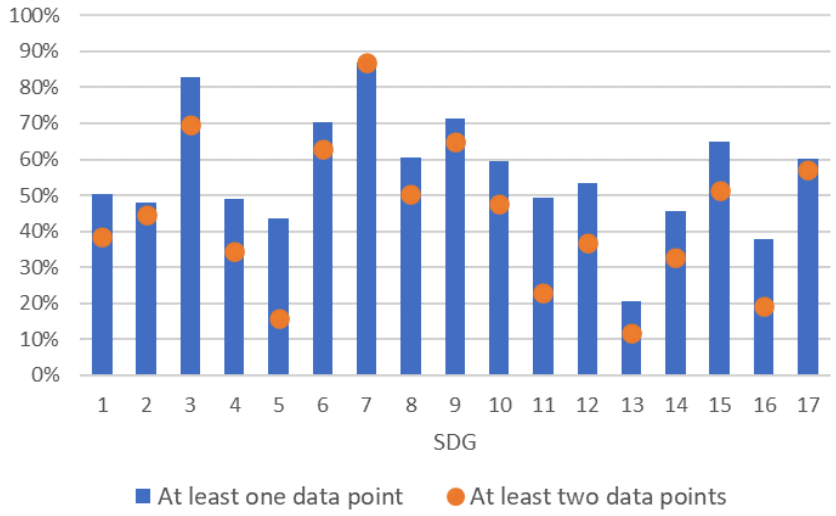
inequalities. In hard-to-reach places, geospatial statistical models relying on satellite imagery and machine-learning algorithms can now estimate population numbers and distributions at a fine spatial scale.

In parallel, more than 40 countries in Europe and parts of Asia have begun using population registers linked with other administrative records (such as on education, employment, taxation, social security and health) to produce census-like data. Some countries have conducted a fully register-based census while others are combining administrative data with partial or full census enumeration, or taking steps to undertake this transition. Progress in using civil registration and vital statistics and identity management information systems for real-time monitoring of population dynamics and vital events remains sparse, however. As such, there is broad consensus amongst National Statistical Offices that field-based population censuses will continue to serve as pillars of population data systems in lower income countries for the 2030 census round. Household survey programmes will also remain important globally in providing development metrics, adding value in understanding changing social norms, and responding to sensitive issues such as sexual and gender-based violence (SGBV).

Multiple applications have demonstrated the potential of digital trace data[2] to complement official statistics, particularly in tracking population mobility. These data also allow measurement when traditional forms of data collection are not feasible, such as during crises, and can reveal trends and patterns before they appear in official data. The proprietary nature of digital trace data, however, seriously limits access for statistical purposes. Bias in representativeness, lack of consent for the reuse of personal data and risks to privacy remain unresolved concerns.

Despite significant progress, gaps persist in population data availability, quality, accessibility and use. Only half or fewer countries have data to monitor SDG indicators on poverty, nutrition, education and gender (see figure 1).

Figure 1 - Percentage of countries with SDG indicator data since 2015



Source: United Nations, Global SDG Indicator Database, accessed February 2024.

[2] Digital trace data are generated by digital technologies such as social media platforms, mobile apps and web search engines, and by off-line information that is digitally captured (e.g., smart electricity meters).

This is particularly the case for data on key populations and some of the hardest to reach groups, such as migrants and refugees, people with non-binary gender identities, persons with disabilities, Afrodescendent people and indigenous groups.

Data gaps result from persistent investment shortfalls and structural deficiencies in national data ecosystems, including the lack of a fully funded national statistical plan; reliance on sub-optimal external donor support; persistent limitations in data accessibility and openness; inadequate technical capacity in both “old” statistical and “new” data science skills; a disconnect between humanitarian and development data; and the lack of effective data governance frameworks for data sharing and protection.

Population data amidst global megatrends

Technological innovations will no doubt continue to shape the data systems of the future. The potential for AI to automate data processing and analysis, identify patterns and correlations in large data sets, and provide predictive analytics is already on the horizon. Yet challenges must be addressed, such as the need for high-quality ground data for model calibration, and the risk of algorithmic biases. While AI is a potential game changer, the misperception that it will soon be the standalone solution for all data needs must be carefully managed.

Intersecting and mutually reinforcing global megatrends are also creating new data needs, for example, on the impact of the climate crisis on livelihoods, reproductive health and the most marginalized subpopulations; on the demand for assisted reproductive technologies; on changing social norms in low-fertility contexts; and on the determinants of healthy ageing over the life course. In a world of frequent and protracted humanitarian crises, more timely and representative data on forced displacement are imperative to guide not only short-term humanitarian responses but also long-term, equitable solutions.

Pathways towards future-ready population data systems

Given different data histories and information systems, there is no one-size-fits-all solution for the modernization and optimization of population data applicable to all national contexts. Nevertheless, key pathways can be identified based on existing best practices and informed expectations about the factors that will continue to shape the demand and supply of population data.

The transition to integrated population data systems

Data integration across censuses, surveys, administrative sources and other types of data is a powerful means to improving the quality and cost-effectiveness of data generation, bridge the humanitarian and development data divide and respond to population and health needs during crises. Integrated population data systems enable governments to locate and address inequalities for achieving the goal of "leaving no one behind", and with the inclusion of gender statistics, they provide governments with important means to monitor and address gender equality and emerging health needs. Key strategies for data integration include a stepwise and context-specific transition to a

register-based population data system, with gender-sensitive and human-centered civil registration and vital statistics systems at the core; enhancing integration of geospatial and statistical data to enable the location of populations burdened by multiple concurring forms of disadvantage, including climate events and other humanitarian crises; and scaling up the use of *non-traditional data sources*, such as mobile phone records and social media, particularly for capturing timely and granular information on migration and different patterns of mobility.

A rights-based data governance architecture

Rights-based data governance and management systems need urgent attention to establish requisite standards that balance open access, data confidentiality and ownership. Where population and health data are being integrated, further elaboration of data governance and protection mechanisms must guard against "function creep", ensure informed consent for data reuse and mitigate risks of misuse amplified by AI-powered algorithms. This is important to reconcile the widening gap between official statistics and the infinitely larger universe of population data produced by the public and private sectors, and to redefine the role and accountability of national statistical offices as reference custodians and stewards of the quality of the data used by society.

Sustaining financing for data and statistics

Long term development benefits can be achieved through effective *domestic financing strategies* that embed data capacity strengthening and resource allocation within sectoral budgets, with the potential to enhance data generation across demographic, social and health domains. While investments in registry data systems should be prioritized, sustained *resource allocations for population censuses and household surveys* in lower-income countries – in combination with cost-optimization approaches for data collection enabled by new technologies – remain important to avoid a widening of the data divide between the Global North and South.

Addressing inequalities in statistical capacity and data use

Strategic data investments in LMICs should address both institutional and skill capacity gaps across the data value chain. Countries can build data competencies at the intersection of population, development and environmental issues, including through enhanced training for young data professionals, and investment in centres of excellence at the forefront of data innovation in the Global South, ensuring that national data are based on local knowledge and priorities. The integration of data literacy into education curricula at various levels can promote a culture of data-driven decision-making across society, in policymaking, programme implementation and advocacy.