



ABOUT

The *Data Bulletin: Informing a Global Compact for Migration* series aims to summarize in an accurate and accessible fashion existing evidence on migration to support the discussions and any follow-up activities of a global compact for safe, orderly and regular migration.

As part of the project “Support to IOM for the Global Compact for Safe, Orderly and Regular Migration,” funded by the European Union, Data Bulletins outline the strengths and limitations of relevant migration data, and highlight innovative data practices which are pertinent to the global compact for migration. Data Bulletins reflect the collaborative nature of this process by including relevant contributions from different parts of IOM as well as other agencies and migration experts.

ADVISORY BOARD

Marla Asis (Scalabrini Migration Centre, Manila)

Alexandra Bilak (Internal Displacement Monitoring Centre, Geneva)

Marcello Carammia (European Asylum and Support Office, Malta)

Andrew Geddes (Migration Policy Centre, European University Institute, Florence)

Yosr A. A. Ibrahim (Head of Sustainable Development Unit, CAPMAS, Cairo)

Diego Iturralde (Statistics South Africa)

Juan Manuel Jiménez (Organization of American States, Washington, D.C.)

Manuel Orozco (Inter-American Dialogue, Washington, D.C.)

Mustafa Hakki Özel (ILO, Geneva)

Jorge Martinez Pizarro (CELADE, CEPAL, Ecuador)

Sonia Plaza (World Bank, Washington, D.C.)

Jon Simmons (UK Home Office, London)

Ann Singleton (University of Bristol / IOM's GMDAC, Berlin)

Keiko Osaki Tomita (UN Statistics Division, New York)

Alessandra Zampieri (Joint Research Centre, European Commission, Ispra)

MANAGING DIRECTOR

Frank Laczo (IOM's GMDAC, Berlin)

Big Data and Migration

How data innovation can serve migration policymaking¹

Migration has become one of the most challenging issues confronting policymakers around the world. The growing complexity of internal and cross-border human mobility has highlighted the need for reliable and timely data to inform migration policy development and humanitarian assistance – a need that traditional statistical systems are often not well-equipped to meet.

Migration data from traditional sources – e.g. national population censuses, sample surveys and administrative sources – are characterized by significant gaps in quantity and quality, despite increasing efforts from national governments and the international community to address these.² National population censuses are conducted infrequently and cannot provide timely information on migration; household surveys are costly and may not always include migrants, particularly if they are undocumented. Administrative records (e.g. issuance of residence and work permits) are not analysed and disseminated systematically in many countries.³

Such limitations create important gaps in the information needed to inform sensible and effective migration policies, and to monitor progress towards the achievement of migration-relevant targets in the SDGs; when migrants are absent from the data, it becomes difficult, if not impossible, to assess the extent to which they are ‘left behind’. The adoption of a global compact for safe, orderly and regular migration in December 2018 will also imply greater demands for data to help make migration safer and more orderly.

Meanwhile, most data today are not collected by national statistical offices but by private companies or international agencies. Estimates suggest that about 90 per cent of the world's data were generated in the last two years.⁴ Technological innovations and the reduction in the cost of digital devices mean that vast amounts of data generated

¹ This bulletin is largely based on the outcome document from the workshop *Big Data and alternative data sources on migration: from case-studies to policy support*, jointly organized by the European Commission's Knowledge Centre on Migration and Demography (KCMD) and IOM's Global Migration Data Analysis Centre (GMDAC) in Ispra, Italy on 30 November 2017. For more details about the event, see <https://bluehub.jrc.ec.europa.eu/bigdata4migration>.

² See, for instance, IFMS website, <http://www.oecd.org/migration/forum-migration-statistics/>, and the Migration Data Portal page on data gaps and availability, <https://migrationdataportal.org/themes/migration-data-sources>.

³ See Data Bulletin Issue no. 4, January 2018 at <http://gmdac.iom.int/data-bulletin-improving-data-for-safe-orderly-and-regular-migration>.

⁴ UN Data Revolution Group (2014) *A World that Counts: Mobilising the Data Revolution for Sustainable Development*, available at <http://www.undatarevolution.org/report/>.



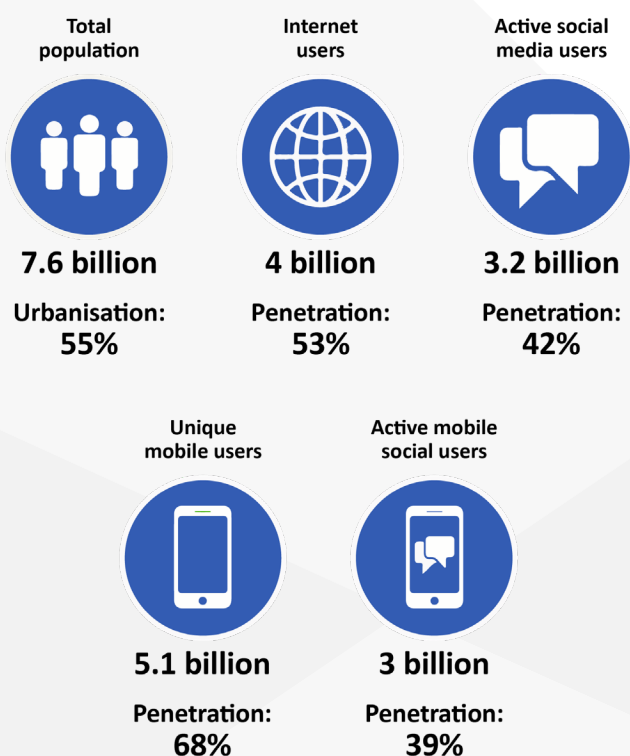
through use of mobile phones, internet-based platforms and other digital devices are now collected in real time, at very little cost. Sensible and responsible analyses of these “digital crumbs” have the potential to offer important insights into societal phenomena, including migration. However, a series of issues, ranging from access and analytical difficulties to privacy and security risks, mean that such vast potential remains largely untapped.

THE POTENTIAL

A growing number of studies and experiments combining traditional and new data sources and methods are bringing the potential of data innovation to the attention of the international migration community. Existing evidence shows that some types of big data sources are particularly promising for the study of migration-related phenomena. These include mobile phone call detail records (CDR); internet activity – such as google searches, repeated logins to the same website, and e-mails sent; online media content; and geo-referenced social media activity, which can be obtained via advertising platforms offered by social media companies to help advertisers target specific audiences.

Digital around the world in 2018

Key statistical indicators for the world's internet, mobile and social media users



Infographic based on Hootsuite and We Are Social, 2018. For sources and further information please see original at: <https://wearesocial.com/uk/blog/2018/01/global-digital-report-2018>

Data obtained from the Facebook advertising platform,⁵ at little or no cost, can provide information on the number of users Facebook refers to as “expats” – those stating to be living in a country other than their self-reported “home country” – in a specific country or globally at a certain point in time.⁶ This means that Facebook can potentially be used as a “real-time census” to estimate the stock of “expats” in a country. According to Facebook data, there would be approximately 273 million “expats” globally (monthly active users), which is not too far from the estimate of 258 million international migrants worldwide by mid-2017 by the UN Department of Economic and Social Affairs (UNDESA).⁷

Geo-located social media activity can also be used to infer patterns of international migration flows, disaggregated by sex, age and other – usually self-reported – characteristics, such as skill levels or sector of occupation. For instance, data from LinkedIn – which counts more than 530 million users globally – allow for a digital mapping of the workforce in certain regions or countries, including migrants and their occupational profiles. LinkedIn data on self-reported changes in job positions can be analysed to estimate movements of highly skilled migrants, particularly in locations and sectors where penetration rates of this social media platform are relatively high.⁸ Twitter data has also been used to compare internal and international migration patterns.⁹ Social media content, for instance from Twitter, may also be used to study public sentiment on migration, and analyse how views on social media can become polarized and self-referential.¹⁰

The combination of independent data sources with different levels of precision with respect to time and location seems particularly promising. For instance, call records from mobile phones combined with satellite data can be used to map movements between cross-

⁵ See <https://developers.facebook.com/docs/marketing-api/targeting-specs/>.

⁶ Zagheni, E., Weber, I. and Gummadi, K. (2017) Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants, available at http://www.zagheni.net/uploads/3/4/4/7/34477700/zagheni_weber_gummadi_2017_accepted_version.pdf.

⁷ UNDESA (2017) The International Migration Report 2017 (Highlights), see <https://www.un.org/development/desa/publications/international-migration-report-2017.html>. Facebook counts about 2 billion users, representing around 30% of the world's population. Penetration rates are much lower in Asia and Africa compared to the rest of the world (only 4% in China, for instance), though such rates may be higher among migrants relative to non-migrants (<https://www.statista.com/statistics/553712/facebook-penetration-in-china/>). The number of fake or double accounts could also be a possible source of error. Also, while UNDESA estimates mostly refer to the foreign-born population globally, Facebook counts as “expats” those who state to be living in a country other than their self-reported “home country.”

⁸ State, B. et al. (2014) Migration of Professionals to the US: Evidence from LinkedIn Data, see http://zagheni.net/uploads/3/1/7/9/3179747/migration_professionals_linkedin.pdf.

⁹ Zagheni, E., Kiran, V.R. and State, B. (2014) Inferring International and Internal Migration Patterns from Twitter Data, see <https://ingmarweber.de/wp-content/uploads/2014/02/Inferring-International-and-Internal-Migration-Patterns-from-Twitter-Data.pdf>.

¹⁰ Presentation by F. Natale at KCMD-GMDAC workshop *Big Data and alternative data sources on migration: from case-studies to policy support*, see fn. 1.

The potential of big data and innovative data sources

Filling the gaps of traditional data sources on migration (censuses, surveys and administrative sources) and contributing to improved understanding of various aspects of migration. A number of studies have shown such potential – but there are significant challenges too.



© IOM's GMDAC 2017 www.migrationdataportal.org

communities.¹¹ While CDR data are generally more helpful to track internal migration patterns – including post-disaster displacement¹² – they could potentially be used to study international migration across neighbouring countries, complementing traditional data, such as administrative records. Also, details of international calls, coupled with traditional statistics from national censuses can contribute to determining patterns of migrant integration and residential segregation.¹³ CDR can also be used to identify ‘transnationals’ – people living and working in more than one country – whom are usually hard to track through traditional data sources.¹⁴

In sum, the main strengths of new data sources, or the combination of new and traditional sources, lie especially in a) their wide coverage, as these data refer to all users of mobile devices or internet services, and therefore vast segments of the population (even those who may be hard to reach);¹⁵ b) their timeliness, as data are collected in

virtually real-time; c) the frequency at which information can be updated; d) the high spatial resolution of big data; e) the relatively low cost of obtaining such data – depending on the willingness of data providers to share them – as they are automatically collected and stored in real time; and f) their potential to generate insights on temporary, circular and seasonal patterns of migration, which are difficult to capture through traditional sources.

THE CHALLENGES

The opportunities offered by new data sources to improve understanding of migration-related aspects are mirrored by numerous challenges: a) there are privacy, data protection and ethical issues in using data automatically generated by individuals of mobile phones and internet-based services, and human rights concerns due to the risks of using such data for surveillance or other incompatible purposes compared to the purposes for which they were collected; b) the volume, complexity and ‘noisiness’ of the data imply analytical and methodological difficulties in extracting meaningful insights; c) big data sources reflect behavioural patterns of users of mobile phones and internet platforms, who may not be representative of the population at large (‘selection bias’), so methodological adjustments are needed to address this bias; d) self-reported information on social media may not always be reliable and the potential presence of fake or double social media accounts also need to be considered; e) access will need to be clearly regulated so that data continuity is ensured and any privacy concerns addressed; f) there are difficulties in applying statistical definitions of migration, such as the UN-recommended definition

¹¹ Presentation by A. Sorichetta at KCMD-GMDAC workshop *Big Data and alternative data sources on migration: from case-studies to policy support*, see fn. 1.

¹² See, for instance, Flowminder "Mobile phone data to understand climate change and migration patterns in Bangladesh," 2016. Available from <http://www.flowminder.org/case-studies/mobile-phone-data-to-understand-climate-change-and-migration-patterns-in-bangladesh>.

¹³ Presentation by F. Natale at KCMD-GMDAC workshop *Big Data and alternative data sources on migration: from case-studies to policy support*, see fn. 1.

¹⁴ Ahas, R., Silim, S. and Tiru, M. "Tracking trans-nationalism with mobile telephone data," (Foundation Estonian Cooperation Assembly, 2017). Available from <https://inimareng.ee/en/open-to-the-world/tracking-trans-nationalism-with-mobile-telephone-data/>.

¹⁵ Coverage varies depending on internet, mobile and social media penetration rates, which differ by region and individuals' characteristics. For instance, the share of internet users in the total population in Africa is 34 per cent, compared to 48 per cent in the Asia-Pacific region, 73 per cent in the Americas and 80 per cent in Europe. Only 15 per cent of the total African population are active social media users, compared to 42 per cent in Asia-Pacific, 53 per cent in Europe and 64 per cent in the Americas (for all sources, see <https://wearesocial.com/blog/2018/01/global-digital-report-2018>).

of an international migrant, given information on length of residence in the host country is mostly hard to obtain from big data sources.

These challenges generate some scepticism and uncertainty among statisticians and policymakers on the feasibility of using big data sources to address gaps in migration statistics and inform migration-related policies.

REALIZING THE POTENTIAL

To overcome the challenges and systematize use of big data sources for migration research and policymaking, investments are needed in the following areas:

1. The establishment of an **adequate regulatory and legislative framework** for the collection, analysis, and sharing of big data; an **international dialogue** between regulators, big data users and providers should be the starting point.
2. The **upgrading of infrastructure and security systems** at the **national level**.
3. The creation of incentives towards the **development of new private-public partnerships** for data exchange and collaborations, or “data collaboratives”.¹⁶
4. More **research** on ways to capitalize on innovative data sources in the field of migration, and systematic ways to take stock of existing applications and share existing knowledge.

¹⁶ See <http://datacollaboratives.org/> and presentation by S. Verhulst at KCMD-GMDAC workshop *Big Data and alternative data sources on migration: from case-studies to policy support*, see fn. 1.

As a way to facilitate these investments, the European Commission's KCMD and IOM's GMDAC are planning to create a Big Data for Migration Alliance (BD4M) to advance discussions on how to harness the potential of big data sources for the analysis of migration and its relevance for policymaking, while ensuring the ethical use of data and the protection of individuals' privacy.

As conveners of BD4M, the KCMD and IOM's GMDAC would welcome the participation of representatives from international organizations and NGOs, members of national statistical offices, private sector representatives, researchers and data scientists interested in contributing, in various capacities, to realizing the potential of big data to complement traditional data sources on migration.

The BD4M would be global in scope, and would aim to make progress in the area of big data and migration through 3 main areas of work:

1. **Awareness-raising and knowledge-sharing:** promoting sharing of knowledge on data innovation in the field of migration.
2. **Capacity-building:** providing technical support to countries interested in using new data sources to complement migration statistics.
3. **Policy-oriented research:** developing research projects to test new big data applications to study migration-related patterns.

The BD4M would also encourage the creation of a network of “data stewards” across private and public institutions, to facilitate exchanges of information and good practices on how to leverage big data for migration analysis, with due consideration of privacy and ethical concerns.

IOM's GMDAC

In response to growing calls for better data on migration, and better use and presentation of migration data, IOM has created the Global Migration Data Analysis Centre (GMDAC).

Located in Berlin, Germany, the Centre aims to provide authoritative and timely analysis of data on global migration issues as a global hub for data and statistics on migration.

IOM's Global Migration Data Analysis Centre (GMDAC)

Taubenstr. 20-22 | 10117 Berlin, Germany

Tel.: +49 30 278 778 21

gmdac.iom.int

migrationdataportal.org

Contact

Marzia Rango

IOM's GMDAC

mrango@iom.int

Michele Vespe

European Commission, Joint Research Centre

michele.vespe@europa.ec.eu

Denis Kierans

IOM's GMDAC

dkierans@iom.int

The opinions expressed in this briefing are those of the author and do not necessarily reflect the views of the International Organization for Migration (IOM). The designations employed and the presentation of material throughout the briefing do not imply the expression of any opinion whatsoever on the part of IOM concerning the legal status of any country, territory, city or area, or of its authorities, or concerning its frontiers and boundaries.

This publication has been produced with the assistance of the European Union. The contents of this publication are the sole responsibility of its authors and can in no way be taken to reflect the views of the European Union.