



BIG DATA AT THE CROSSROADS: SEIZING THE POTENTIAL OF BIG DATA TO GUIDE THE FUTURE OF EU MIGRATION POLICY

Matías Ibáñez Sales

Programme Officer, Department of Sustainable Development and Regional Integration, European Institute of the Mediterranean (IEMed)

Introduction

“Without good data, we’re flying blind. If you can’t see it, you can’t solve it,” stated former United Nations (UN) Secretary-General Kofi Annan in 2018. The growing, and sometimes unexpected, complexity of transnational human mobility has emphasised the need for reliable data to facilitate timely evidence-based policy processes on migration phenomena.

Migration data acquired from traditional sources such as national censuses, sample surveys or administrative processes presents persistent gaps in the absence of real time information, poor frequency of update and low comparability across countries, despite the progress made in recent decades. The disruptive time gap between the collection of data and the real availability of such information makes policy processes slow down without the chance to respond to the real needs of the target groups in a timely manner. Moreover, migrants with an irregular status are often excluded from migration statistics of the transit and host countries. If migrants cannot be counted, tracked or monitored, they will therefore always be absent from decision planning and the implementation of targeted migration and integration policies (Rango & Vespe, 2017).

Big data offers the real possibility of starting to address this issue and closing the existing data gap. The term big data includes large volumes of high velocity, complex and variable data, which requires advanced technologies to capture, store, distribute and analyse such information. Current technological innovations in data analysis have advanced over the last few years and now permit these *digital footprints* to be effectively transformed into actionable information that informs decisions (De Mauro et al., 2016), despite some technical and methodological challenges that are still present today (Ziesche, 2017).

According to Gandomi & Haider (2014), today 95% of big data is originated automatically by users of mobile phones, social media, online platforms, search engines and applications, including geo-localisation sensors. Consequently, big data enables complex data to be captured with high spatial resolution; for instance, a sample of the totality of migrants using mobile phones or Internet-based applications (Hilbert, 2014). Moreover, it provides constant flows of update making this data accessible in real time and at a much lower cost than it would involve to gather the same information through administrative processes (United Nations Global Pulse, 2016; Ziesche, 2017).

In this context, big data stands out as a key tool to complement traditional sources of migration data by providing evidence-based and timely information on migration phenomena that would otherwise be very difficult to analyse through traditional methods of data collection; for instance, on irregular migration flows, intraregional mobility or inclusion outcomes (Rango & Vespe, 2017).

The aim of this paper is to assess the prospects and opportunities of using big data for the analysis of migration-related phenomena that are not easily observable through traditional means and the feasibility of better informing migration and integration policies. Concrete examples of innovative applications of big data sources in the field of migration will show its potential to complement traditional migration data sources and its relevance to opportunely inform day-to-day migration policy decision and implementation processes. In addition, this paper will shed light on some of the existing challenges that are still associated with big data uses. Finally, a series of recommendations addressed to the European Commission (EC) will reflect on possible action points to seize the potential of big data on migration policy while mitigating some of its current risks and challenges.

Migrants and refugees constantly generate digital footprints

Nowadays, migrants and refugees are increasingly using mobile devices and digital interactions on social media to communicate with relatives and to send them news and pictures. In addition, smartphones are being used by migrants to facilitate their movements across borders, including to contact human smugglers, to stay up to date with weather predictions, to get last minute information in transit and destination countries or to send and receive international remittances. For instance, it has also been documented that groups of migrants crossing the Mediterranean Sea to Europe on rubber boats used their mobile phones on some occasions to call national coastguards and share the exact location of the vessel before the occurrence of a shipwreck (Eide, 2020).

According to a recent Europol report (2018), there has been an exponential growth in the use of social media by refugees and migrants arriving in Europe in the last few years. Even the United Nations High Commissioner for Refugees (UNHCR) has concluded that smartphones and social media apps are now a key tool for migrants, who spend up to a third of their total budget on staying connected (2017).

In summary, there are three main types of big data sources that are highly relevant to complement traditional migration data sources: communication services (calls records and texts messages), geo-localised activity as well as Internet-based exchanges (social media activity, online searching preferences, and online money transfers).

Concrete examples of the potential of big data applied to migration and integration policy

Today, a growing number of studies, pilot projects and experiments from the academic, private and public sphere are combining traditional data sources with innovative collecting methods. In the following paragraphs a number of selected applications of big data on migration policy as well as its opportunities will be showcased.

To begin with, big data in combination with machine learning technology could be used to produce advanced migrant admission and integration systems at the national level. As an example, big data could help create migrants' profiles to better assess the review of their visa applications and complement information submitted to immigration authorities. Nowadays, for instance, the United States (US) has started to require social media information on those migrants planning to enter the country (OECD, 2020).

At the same time, the combination of big data with artificial intelligence allows the prediction of potential future integration outcomes based on social and economic variables. As successful examples in this field have recently shown, countries with access to big data and non-observable characteristics could predict future labour market and socio-economic integration outcomes of newcomers based on the performance of past migrants (OECD, 2020). For instance, the Immigration Policy Lab (IPL) at Stanford University and ETH Zurich developed a big data-driven algorithm in 2018 to optimise the process by which refugees are assigned to certain locations within a resettlement country (Bansak et al., 2018). Today, regular resettlement methods used in many European countries are based on the specific allocation of asylum seekers across different cities, based on the actual space available in each region or the political will to receive them. However, this innovative system uses historical data on former refugees' characteristics (age, gender, origin, skills, etc.) and their labour market performance in each region to calculate the best probability of new refugees finding a job at each resettlement location. The project started to be implemented as a pilot test in Switzerland in 2018 and initial results suggest that if the algorithm had been applied earlier, refugee employability rates could have experienced an increase up to a 70% in the preceding years (Bansak et al., 2018).

Second, geo-localised social media activity as well as self-reported information based on popular apps, such as Facebook or Twitter, can be used to identify current and ongoing international migration flows and stocks that are currently hard to capture through traditional data means, as well as to inform decision-makers about potential arrivals in the short and medium term.

Some studies conducted by the Qatar Computing Research Institute have helped calculate an estimate number of worldwide expats based on Facebook users self-reported information to be living in a different country from their place of origin, and thus helping complement international statistics on this issue (Mejova et al. 2018). In the same line, the EC's Directorate-General for Employment, Social Affairs and Inclusion commissioned RAND Europe to investigate big data's potential use for complementing the European Union (EU)'s official stats on stocks of European citizens on the move and EU labour mobility flows (Gendronneau et al., 2019). By collecting geo-referenced social media information and data based on the approximate number of users within the EU that self-reported on Facebook to be living in another member state, the research carried out presents real-time estimations on changes in EU mobility that will help complement traditional migration statistics from Eurostat and national census, which is vital to inform further EU and national policy decisions (Gendronneau et al., 2019). Moreover, LinkedIn has been successfully tested for digital mapping of international workers in certain regions, while estimating likely movements of high- and low-skilled migrants according to changes in their job status (Rango & Vespe, 2017).

Third, some pilot experiments have recently shown that big data, especially social media activity, call records and satellite data could be a potential ally for risk assessment and disaster management. Particularly, big data can be used to identify in advance potential large-scale migratory movements and trace cross-border migration in the event of natural disasters, conflict outbreaks or economic downturns (Ravenna Sohst & Tjaden, 2020). Numerous pieces of research have shown (Böhme et al., 2020; Napierala et al., 2021) that obtaining timely evidence-based information on potential migratory movements, sometimes up to months in advance, is of the utmost importance to develop early-warning systems and set-up the necessary reception mechanisms in transit and receiving countries.

In this context, a study conducted by the Pew Research Center successfully measured migration intentions from Turkey to Greece based on big data aggregated from Google searches in Arabic for words such as “Greece” or “crossing” between the years 2015 and 2016 (Connor, 2017). The results of the study have shown a high relationship with the final number of asylum seekers arriving in Greece reported by UNHCR in 2016.

Similarly, the European Asylum Support Office (EASO) combines geo-localised media activity with historical migration data to estimate potential scenarios in neighbouring countries that could lead to an increase of asylum-related migration towards the EU. EASO’s so-called *Early Warning and Forecasting System* is based on the analysis of disruptive situations such as conflicts, political instability and natural-based events from a media database called Gdelt, which gathers media reports and news from all over the world (Napierala et al., 2021; Melachrinou et al., 2020). Complementarily, an algorithm-based machine referred to as the “Push Factor Index” (PFI) weighs and categorises such information to create individualised country-level indexes, develops potential scenarios of large and forced migration inflows for a period of up to three weeks in advance, and shares in real time this information with member states, which is aimed at mobilising the operational set-up of early-warning mechanisms and asylum-claim systems (Melachrinou et al., 2020).

The perils of big data: a series of challenges for the future

Some policy-makers, researchers and migration stakeholders remain, however, sceptical about using big data as a means of informing migration governance and integration policies, given the existing challenges. These challenges nowadays generate concern and uncertainty on the feasibility, accuracy and ethics of big data use to address gaps in migration statistics.

Since the vast majority of big data is generated automatically by mobile and Internet users, often without informed consent, its use may lead to violations on the grounds of privacy and data protection. Issues surrounding privacy and confidentiality continue to pose risks and challenges to migrants as effective mechanisms to ensure governments and corporations are held accountable for breaching migrants’ data privacy still do not exist (Bircan & Korkmaz, 2021). This situation also brings to the table concerns on public-private cooperation and engagement as governments may become dependent on data collection, storage and analysis by outsourcing these services to corporations, which can result in a mass data commercialisation (Bircan & Korkmaz, 2021).

Ethical debates could also arise about applying the same nationals’ data privacy rights to migrants, as public opinion may oppose collecting big data based on grounds of personal information reserved to the private sphere, such as sexual identity or ideology (OECD, 2020). At the same time, migrants’ human rights could be at stake if states use the forecasting information on migratory movements to close their national borders and further criminalise incoming asylum seekers by preventing them from accessing international protection

(Beduschi, 2017). Following this same line of thought, information collected through phone call records and geo-localised sensors of migrants' smartphones could lead to a serious breach of civil liberties and fundamental freedoms if it is used for digital surveillance purposes (Taylor, 2015).

Moreover, big data use also brings methodological concerns as it implies a twofold digital divide between countries, as some countries possess better equipment to collect and analyse big data in real time, and between human beings (Ziesche, 2017). Since digital migrant users are not a truthful representation of the whole migrant population, big data sources are often criticised for not being immediately reliable or completely accurate but require careful interpretation and bias correction. Moreover, some authors reflect that big data could bring a widening gap between digital information-rich and poor countries, and the data sample per se would not be reliable for policy-making purposes (IEAG, 2014).

Last but not least, there are still significant technical and analytical challenges on big data use due to the difficulties in accessing raw data mostly held by private companies and problems in aggregating value to complex amounts of digital data (Ziesche, 2017). Insufficient infrastructures to store massive amounts of data and lack of quality of cybersecurity mechanisms in the public sector to prevent hacking of systems and leaks of migrants' information are also worth mentioning.

Conclusions and recommendations

Having real time, accurate and high-quality data on migration patterns and phenomena is the first step towards the implementation of concrete policy solutions tailored to the right target groups. The Global Compact for Safe, Orderly and Regular Migration (2018) calls on states in its first objective to improve the collection, use and dissemination of accurate and disaggregated data as a basis for evidence-based policies. More precisely, it specifically refers to the need to address data gaps by leveraging new data sources, including big data. Progress on this matter has been made and a series of initiatives and projects were recently developed by the UN to assess the potential benefits of big data on complementing other traditional data sources for monitoring SDG indicators, for instance the UN Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG), the UN Global Pulse, the UN Global Working Group on Big Data for Official Statistics, and the UN Data Innovation Lab.

The new Agenda for the Mediterranean, recently launched by the EC in February 2021, presents green and digital transitions as flagship priorities for the upcoming years in North-South cooperation. However, its point number four related to "*Migration and Mobility*" does not specifically mention how the EU could make use of technological advances to enhance migration management and strengthen cooperation with the southern neighbourhood.

1. In this respect, European member states and the EC should seize the potential of big data and new technologies to complement traditional migration data sources and to better inform migration policy processes, while mitigating current risks and challenges.

In this line, existing integration programmes at the national and sub-national level should be updated and reformulated considering new available technologies, which in combination with big data can help positively improve the programme's compliance, monitoring and efficiency as well as migrants' selection and labour market performance. Moreover, innovative applications of big data in the field of migration management and risk assessment should be strategically endorsed to help improve European migration

governance and foster innovative early-warning mechanisms. Complementarily, the EU should seize the potential of big data to complement the EU's and member states' migration statistics on migration-related phenomena that is hard to capture through traditional means.

In all these processes, new migration mechanisms should be transparent and avoid basing migration decisions exclusively on big data and artificial intelligence. A possible mitigation measure would be to ensure that a law enforcement official is always in charge of the final decision. Moreover, public-private cooperation towards the development of synergic frameworks should be further assessed. To avoid misuses and breaches of data privacy, public ownership service providers should be predominantly considered for complex selection mechanisms while public-private mechanisms should be encouraged for non-sensitive migratory issues.

2. Moreover, further EC efforts should be made towards the creation of adequate and timely regulatory and legislative frameworks for the collection, use, analysis and sharing of big data in the field of migration.

New regulations adopted at the European level should particularly ensure the protection of migrants' personal data and privacy and prevent public and private holders from using big data against human rights and fundamental freedoms by developing adequate accountability mechanisms. Because as former UN Secretary-General Kofi Annan reminded us in 2015, "*without the right policy framework ensuring that big data is not abused, our privacy, indeed our liberty, will be in danger.*" The need to set up agreed core principles concerning legal, privacy and ethical standards remains crucial to guide the future of the big data revolution while ensuring a migrants' rights-approach, always in conjunction with international developments in big data regulation. Moreover, the EU should seek to promote cooperation with international regulatory and policy-oriented bodies, such as the UN and the Organisation for Economic Co-operation and Development (OECD), which are already working on the nexus of migration, sustainable development and big data.

3. The EC should further enhance policy discussions and multi-stakeholder dialogues on how to seize the potential of big data applied to migration governance.

In this line, awareness-raising initiatives and knowledge-sharing activities in the form of forums, dialogues and gatherings with migration actors and experts from the whole data ecosystem should be further promoted. In this context, the EC should continue taking an active part in the creation of new platforms and the promotion of existing networks, both at the international and regional level, that aim to generate knowledge on big data and migration, sharing best practices and assessing new developments in the field of big data. For instance, the Big Data for Migration Alliance (BD4M) was jointly created by the International Organization for Migration (IOM) and the EC in 2018 to shed some light on the interlinkages between these two fields. By sharing up-to-date knowledge on big data applications, producing policy-oriented recommendations and high-quality research, and enhancing capacity-building activities among European member states, the alliance should take an active role in helping inform and advance the policy and regulatory processes in the EU. A second example is the current Eurostat's Task Force on Big Data, which has a range of pilot projects in different fields. This task force could be expanded by including programmes on big data

use applied to migration governance in order to produce smart statistics that help guide the future of migration policies.

In conclusion, the use of big data offers a wide range of benefits for improving the design and implementation of EU migration policy, although these opportunities do not come without risks. Big data brings to the table unprecedented possibilities for informing policy decisions and transforming both our societies and the life of migrants. Because without high-quality and up-to-date data sources providing the right information about migration phenomena in real time, decision and policy-making processes become almost impossible. Thus, the need to foster the debate and discussions among all types of migration stakeholders and actors, statisticians and data scientists will remain crucial in the upcoming years to help policy-makers better understand the scope of the big data revolution and guide them in further regulatory and policy implementation processes.

References

- ANNAN, K. (2018, February 2018). Data can help to end malnutrition across Africa. *Nature*. Retrieved from <https://www.nature.com/articles/d41586-018-02386-3#:~:text=The%20data%20indicate%20that%20no,develop%20policies%20and%20track%20accountability>.
- ANNAN, K. (2015, August 15). [Keynote address] Massachusetts Institute of Technology, Cambridge, MA, United States.
- BANSAK, K., FERWERDA, J., HAINMUELLER, J., DILLON, A., HANGARTNER, D., LAWRENCE, D., & WEINSTEIN, J. (2018). Improving refugee integration through data-driven algorithmic assignment. *Science*, 359 (6373), 325-359.
- BEDUSCHI, A. (2017). The Big Data of International Migration: Opportunities and challenges for states under International Human Rights Law. *Georgetown Journal of International Law*, 49 (4). Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3124199
- BIRCAN, T., & KORKMAZ, E.E. (2021). Big data for whose sake? Governing migration through artificial intelligence. *Humanities Social Sciences Communications*, 8. Retrieved from <https://www.nature.com/articles/s41599-021-00910-x>
- BÖHME, M.H., GRÖGER, A., & STÖHR, T. (2020). Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, 142.
- CONNOR, P. (2017). *The digital footprint of Europe's refugees. Online searches in 2015 and 1016 open window into path, timing of migrant flows from Middle East to Europe*. Pew Research Center. Retrieved from <https://www.pewresearch.org/global/2017/06/08/digital-footprint-of-europes-refugees/>
- DE MAURO, A., GRECO, M., & GRIMALDI, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, 65 (3), 122-135.
- EIDE, E. (2020). Mobile flight: Refugees and the importance of cell phones. *Nordic Journal of Migration Research*, 10 (2), 67–81.
- EUROPEAN UNION AGENCY FOR LAW ENFORCEMENT COOPERATION (EUROPOL) (2018). *Two years of EMSC: Activity report January 2017 – January 2018*. European Migration Smuggling Centre, EUROPOL. Retrieved from: https://www.europol.europa.eu/cms/sites/default/files/documents/two_years_of_emsc_report.pdf
- GANDOMI, A., & HAIDER, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35 (2).
- GENDRONNEAU, C., WIŚNIEWSKI, A., YILDIZ, D., ZAGHENI, E., FIORIO, L., HSIAO, Y., STEPANEK, M., WEBER, I., ABEL, G., & HOORENS., S. (2019). *Measuring Labour Mobility and Migration Using Big Data*. Directorate-General for Employment, Social Affairs and Inclusion (DG EMPL). European Commission. Publications Office of the European Union.
- HILBERT, M. (2015). Big Data for development: A review of promises and challenges. *Development Policy Review*, 34 (1)

INDEPENDENT EXPERT ADVISORY GROUP ON A DATA REVOLUTION FOR SUSTAINABLE DEVELOPMENT (IEAG). (2014). *A world that counts: Mobilising the data revolution for sustainable development*. Data Revolution Group. Retrieved from: <https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>

MELACHRINOS, C., CARAMMIA, M., & WILKIN, T. (2020). Using big data to estimate migration “push factors” from Africa. In P. Fargues & M. Rango (Ed.), *Migration in West and North Africa and across the Mediterranean* (pp. 98-113). International Organization for Migration Retrieved from: <https://publications.iom.int/system/files/pdf/migration-in-west-and-north-africa-and-across-the-mediterranean.pdf>

MEJOVA Y., ARAUJO, M., WEBER, I., & AUPETIT, M.J., (2018). Creating a finegrained digital census using Facebook advertising audiences: The case of Doha. *Qatar Foundation Annual Research Conference*.

NAPIERAŁA, J., HILTON, J., FORSTER, J. J., CARAMMIA, M., & BIJAK, J. (2021). Toward an early warning system for monitoring asylum-related migration flows in Europe. *International Migration Review*, 51 (1).

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). (2020). *Towards 2035: Strategic Foresight - Making Migration and Integration Policies Future Ready*. Retrieved from <https://www.oecd.org/migration/mig/migration-strategic-foresight.pdf>

RAVENNA SOHST, R., & TJADEN, J. (2020). Forecasting migration: A policy guide to common approaches and models. *Migration Policy Practices*. X(4). Retrieved from: <https://publications.iom.int/system/files/pdf/mpp-43.pdf>

TAYLOR, L. (2015). No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *SAGE Journals*, 34(2),319-336.

RANGO, M., & VESPE, M. (2017). *Big Data and alternative data sources on migration: From case-studies to policy support. Summary Report*. European Commission, Joint Research Center.

UNITED NATIONS GLOBAL PULSE (2016). *Integrating Big Data into the monitoring and evaluation of development programmes*. Retrieved from http://unglobalpulse.org/sites/default/files/IntegratingBigData_intoMEDP_web_UNGP.pdf

UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES (UNHCR). (2017). *From a refugee perspective: Discourse of Arabic speaking and Afghan refugees and migrants on social media from March to December 2016*. Retrieved from: <https://www.unhcr.org/publications/brochures/5909af4d4/from-a-refugee-perspective.html> <https://reliefweb.int/sites/reliefweb.int/files/resources/58018.pdf>

ZIESCHE, S. (2017). *Innovative Big Data approaches for capturing and analyzing data to monitor and achieve the SDGs*. United Nations Economic Social Common for Asia Pacific (ESCAP). Retrieved from: <https://www.unescap.org/sites/default/d8files/knowledge-products/Innovative%20Big%20Data%20Approaches%20for%20Capturing%20and%20Analyzing%20Data%20to%20Monitor%20and%20Achieve%20the%20SDGs.pdf>