Chapter 18
New Migration Data: Challenges and Opportunities

Francesco Rampazzo, Marzia Rango, and Ingmar Weber

Abstract Migration is hard to measure due to the complexity of the phenomenon and the limitations of traditional data sources. The Digital Revolution has brought opportunities in terms of new data and new methodologies for migration research. Social scientists have started to leverage data from multiple digital data sources, which have huge potential given their timeliness and wide geographic availability. Novel digital data might help in estimating migrant stocks and flows, infer intentions to migrate, and investigate the integration and cultural assimilation of migrants. Moreover, innovative methodologies can help make sense of new and diverse streams of data. For example, Bayesian methods, natural language processing, high-intensity time series, and computational methods might be relevant to study different aspects of migration. Importantly, researchers should consider the ethical implications of using these data sources, as well as the repercussions of their results.

18.1 Introduction

Migration has become one of the most salient issues confronting policymakers around the world. The historic adoption of the Global Compact for Safe, Orderly and Regular Migration (GCM)—the first-ever intergovernmental agreement on international migration—and the Global Compact for Refugees in December 2018 and the inclusion of migration-related targets in the 2030 Agenda for Sustainable

F. Rampazzo (✉)
Leverhulme Centre for Demographic Science, Department of Sociology, Nuffield College, University of Oxford, Oxford, UK
e-mail: Francesco.rampazzo@demography.ox.ac.uk

M. Rango
UN Operations and Crisis Centre (UNOCC), New York, NY, USA
e-mail: marzia.rango@un.org

I. Weber
Universität des Saarlandes, Saarbrücken, Germany
e-mail: ingmar.weber@uni-saarland.de

E. Bertoni et al. (eds.), *Handbook of Computational Social Science for Policy*,
https://doi.org/10.1007/978-3-031-16624-2_18

Development are a clear testament to this. These frameworks have also provided a renewed push to calls from the international community to improve migration statistics globally. The first of the 23 objectives of the GCM is about improving data for evidence-based policy and a more informed public discourse about migration. As a matter of fact, many countries still struggle to report basic facts and figures about migration, which limits their ability to make informed policy decisions and communicate those to the public, but also limits the ability of researchers to contribute to the production of evidence and knowledge on migration.

Migration is a complex phenomenon to measure. Population changes generally happen slowly as fertility and mortality tend to impact population dynamics gradually. However, a country's population structure might change more rapidly due to migration (Billari, 2022). Migration, and in particular international migration, has become increasingly important in shaping population change, especially in higher-income countries, where fertility is decreasing (Bijak, 2010). The study of migration is affected by many challenges (i.e. availability of data, measurement problems, harmonisation of definitions) (Bilsborrow et al., 1997). Above all, there is a lack of timely and comprehensive data about migrants, combined with the varying measures and definitions of migration used by different countries, which are barriers to accurately estimating international migration (Bijak, 2010; Willekens, 1994, 2019). Despite the best efforts of many researchers and official statistics offices, international migration estimates lack quality due to the limited data available in many countries (Kupiszewska & Nowok, 2008; Poulain et al., 2006; Zlotnik, 1987). Migration is a topic widely discussed in several research fields including demography (Lee, 1966), sociology (Petersen, 1958), political science (Boswell et al., 2011), and economics (Kennan & Walker, 2011). Insufficient availability of quality data on migration can have a high social and political impact, because these inaccuracies might limit the capacity to take evidence-based decisions.

The main data sources used to measure migration are censuses, administrative records, and household surveys, collectively referred to as 'traditional data sources'. These data sources have limitations related to the definition of migrants (i.e. the discrepancy between internationally recommended definition and applied definitions in each country), coverage of the entire migrant population, and the quality of the estimates (especially for admin records) (Azose & Raftery, 2019; Willekens, 2019). Moreover, traditional data on migration are not promptly and regularly available. There might be a gap of several months or even years between the time the data are collected and statistics are released to the public. Timely and granular migration data are needed not only for research purposes but also for informed policy and programmatic decisions related to migration. In times of global crisis, such as the COVID-19 pandemic or the Russian invasion of Ukraine, the need for accurate and timely data becomes particularly urgent, but the capacity to collect data from traditional sources can be significantly reduced (Stielike, 2022).

In the last 25 years, the world has experienced a data revolution (Kashyap, 2021). New data created by human digital interactions increased dramatically in volume, speed, and availability. The data revolution did come not only with the advent of new data sources but also with increased computational power. This, in turn, helped to

create more sophisticated models to study social phenomena such as migration. New 'ready-made' data from digital sources, commonly referred to as 'digital trace data' (Salganik, 2019), have started to be repurposed to answer social science questions.

Cesare et al. (2018) addressed the challenges faced by social scientists when using digital traces. One of the main challenges is related to bias and non-representativeness, as users of social media platforms, for instance, are not representative of the broader population and might not necessarily reveal their true opinions or personal details. Correspondingly, understanding how to measure the bias of these online non-representative sources is critical to infer demographic trends for the wider population (Zagheni & Weber, 2015). Once the biases are quantified, one possible next step is to combine different data sources to extract more information and enhance the existing data. This is an ongoing process in which social scientists have started to combine survey data with digital traces, originally created for marketing, and repurposing them for scientific research (Alexander, Polimis and Zagheni, 2020; Gendronneau et al., 2019; Rampazzo et al., 2021; Zagheni et al., 2017). The idea of repurposing data is not new to the social sciences (Billari & Zagheni, 2017; Sutherland, 1963; Zagheni & Weber, 2015). For example, John Graunt's first Life Table (1662) was in fact a reworking of public health data from the *Bills of Mortality* to infer the size of the population of London at the time (Sutherland, 1963).

New data sources are a gold mine for migration studies because they offer an opportunity to address the lack of information which hinders this field of research. Digital traces (especially social media data) are quick to collect using, for example, Twitter's or Facebook's application programming interface (API)[1] (for a comprehensive overview of digital trace data for migration and mobility, check Bosco et al., 2022). This allows to know in close to real time how many of the users are in a specific location and have recently changed their country of residence or are foreign-born, contributing to 'nowcasting' migration (e.g. monitoring trends almost in real time). However, digital traces are not always available to academics and practitioners, as they are mostly owned by businesses and may not be fully and publicly accessible.

This chapter has two objectives. First, it aims to bring examples of how new data sources and methodologies have been used for studying migration and migrant characteristics. Second, it highlights advantages, limitations, and challenges of digital trace data in migration research.

[1] An API is a kind of middleman between data held by a company and a user requesting this data. While the actual database storing the data is protected and not exposed to the outside world, an API provides a link between the requesting user and the server where the data are stored in a database (Cooksey, 2014; Sloan & Quan-Haase, 2017). To be able to connect to an API, a key authentication is usually needed, which is a long series of letters and numbers that identifies the account querying the API (Cooksey, 2014).

18.2 New Data in Migration Research

As a statistical concept, international migration has been historically characterised by five building blocks:[2] (i) legal nationality, (ii) residence, (iii) place of birth, (iv) time, and (v) purpose of stay (Zlotnik, 1987). As these blocks are complexly entwined with each other, statistical systems use one or a combination of them to gather data on international migrants. The United Nations recommends a definition of international migration which explicitly focuses on residence and time (UN, 1998), defining a migrant as a 'person who moves from their country of usual residence for a period of at least 12 months'. Migrants that stay between 3 and 12 months are considered to be short-term migrants. The intended purpose of the UN's definition of international migrants is to harmonise data sources worldwide. However, current definitions of migrants vary between countries. While they all depend on the time of stay outside of the country of usual residence, definitions applied at the national level differ (i.e. 'minimum duration of stay in the destination country required for the change of residence in the origin country' Kupiszewska and Nowok, 2008, p. 58) (Kupiszewska & Nowok, 2008; Willekens, 1994).

It has been suggested that digital traces can help refine migration theory and modelling. Fiorio et al. (2017) and Fiorio et al. (2021) highlight the potential of using geotagged Twitter data to investigate short-term mobility and long-term migration. Indeed, the definition of an international migrant has become tied up with the increase in the number of individuals living transnational lives (Carling et al., 2021). Digital trace data might help broaden or qualify the distinction between short-term and long-term migrants, adding nuances. However, we need to consider that digital trace data do not follow the same definition as traditional data sources. For example, on Twitter, migrants can be identified through changes in their location over a period of time, while Facebook provides on their Advertising Marketing Platform a variable that can be used to characterise migrants. The Facebook variable is defined as 'People that used to live in country x and now live in country y' (Rampazzo et al., 2021), which refers to the concept of residence and usage of the social media. The Facebook migrant definition does not account for the time aspect, which creates problems when comparing official migration statistics and Facebook estimates. In Zagheni et al. (2017), the description of the Facebook migrant variable was 'Expat from country x', which highlights that the definition behind this variable may be subject to change.

The information on the categorisation of migrant users on social media is limited. In the case of Facebook, the evidence comes from internal and external research. Migrant users might be identified not only through self-declared public information (e.g. 'hometown') but also through inferred information based on their use of the

[2] The UN Expert Group on Migration Statistics is updating and revising concepts and definitions on international migration: https://unstats.un.org/unsd/demographic-social/migration-expert-group/task-forces/TF2-ConceptualFramework-Final.pdf and https://unstats.un.org/unsd/demographic-social/migration-expert-group/task-forces/taskforce-2.

social media (e.g. user's IP address) (US SEC Commision, 2018, 2019, 2020). Spyratos et al. (2018) conducted a survey of 114 Facebook users asking them to check whether they were classified by the Facebook Advertising Platform as migrants. The majority of the non-representative sample was classified correctly as an 'expat' despite not having self-reported country of birth or of previous residence on Facebook. Moreover, Facebook's researchers declared to use 'hometown' as a feature for characterising migrants (Herdağdelen et al., 2016). On Twitter, migrants are typically identified through geo-targeting for research studies. However, the number of geo-tagged tweets is limited: only 2/3% of the tweets are provided with a geo-location (Halford et al., 2018; Leetaru et al., 2013). Fake and duplicate accounts might also be a challenge when studying migrants on social media. For Facebook, the percentages of fake and duplicated accounts are reported every year on the US Securities and Exchange Commission documents and are stable at a 11% duplicate accounts and 5% fake accounts (US SEC Commision, 2018, 2019, 2020). Therefore, possible algorithm changes on the measure provided may affect continuity of data from these sources. Case in point, previous work (Palotti et al., 2020; Rampazzo et al., 2021) identified discontinuities in the Facebook data in March 2019 leading to a drop in the global estimates of the number of migrants active on the platform.

Although migrants are not clearly defined in digital trace data, stock estimates of migrant populations seem to be proportionally comparable to traditional data estimates. Zagheni et al. (2017) showed that Facebook Advertising data and American Community Survey data are highly correlated. Moreover, Facebook Advertising data has proved to be faster in capturing out-migration from Puerto Rico in the aftermath of Hurricane Maria. Alexander et al. (2020) show how Facebook Advertising data allowed to provide monthly estimates of the relocation of Puerto Ricans to mainland USA, and subsequent return migration, which traditional data sources were not able to register. The same result is supported by the use of Twitter data (Martín et al., 2020), as well as by monthly Airline Passenger Traffic data used by the US Census Bureau.[3] Facebook Advertising Platform could also be used to monitor out-migration from a country experiencing political turbulence, such as Venezuela (Palotti et al., 2020). These examples highlight another important feature of digital trace data: their broad geographic availability. These data can be widely available also in contexts of poor traditional statistics (e.g. low- and middle-income countries); for example, the Facebook migrant variable is available for 17 of the 54 African countries (Rampazzo & Weber, 2020).

Facebook Advertising data has also provided insights on migrant integration in Germany and the USA (Dubois et al., 2018; Stewart et al., 2019). Cultural assimilation was studied through the comparison of interests expressed online by the German population and Arabic-speaking migrants in Germany (Dubois et al., 2018). Results shows that Arabic-speaking migrants in Germany are less culturally similar compared to other European migrants in Germany, but the divide is less

[3] https://www.census.gov/library/stories/2020/08/estimating-puerto-rico-population-after-hurricane-maria.html

pronounced for younger and more educated men. Similarly, cultural integration in the USA was investigated through self-reported musical interests between Mexican first- and second-generation migrants and Anglo and African Americans (Stewart et al., 2019). The comparison between self-reported musical interests highlights that education and language spoken (e.g. English versus Spanish) are key characteristics determining assimilation. However, these studies are affected by limitations linked to self-reported information and 'black box' algorithms estimating interests on social media platforms.

Analysis of digital traces can do more than help with estimation of current migration stocks. Non-traditional data sources can also provide insights into migration intentions, migration flows, and more. For example, Google Trends data going back to 2004 has been used to estimate migration intentions and subsequently predict flows to selected destination countries (Böhme et al., 2020). Böhme et al. (2020) complemented Google Trends with survey data to predict migration flows and intentions. Their results are robust, but the authors highlight as a limitation that the predictive power of words chosen might change over time. Moreover, the models had higher performance when focusing on countries where internet usage is high (Böhme et al., 2020).

Wanner (2021) used a similar approach with Google Trends data to study migration flows to Switzerland from France, Italy, Germany, and Spain. They found that Google Trends data can anticipate migration flows to a certain extent when actual migration is decreasing in volume. Avramescu and Wiśniowski (2021) focused on Google Trends searches related to employment and education from Romania directed to the UK, creating a composite indicator in a time series model. They obtained mixed results in terms of predictive power, stressing that knowing the context of the origin and destination countries is important to increase accuracy of the predictions. Despite the challenges, all the authors agree that Google Trends is a powerful source for estimating potential migration.

New opportunities might arise also from consumer data from the retail sector (e.g. from basket analysis). For instance, some studies show how food consumption patterns can shed light on integration aspects (Guidotti et al., 2020; Sîrbu et al., 2021). Moreover, companies such as LinkedIn, Indeed, and Duolingo provide reports on their users that might reflect migration dynamics. LinkedIn[4] and Indeed[5] reports focus on economic migration, providing insights on the international job market, while Duolingo[6] featuring the most studied language per country shows, for example, how Swedish is the most popular language in Sweden or that German is the top language studied in the Balkans.

[4] https://www.ecb.europa.eu/pub/economic-bulletin/articles/2021/html/ecb.ebart202105_02~c429c01d24.en.html#toc4

[5] https://www.hiringlab.org/uk/blog/2021/10/05/foreign-interest-in-driving-jobs-rises-on-visa-announcement/

[6] https://blog.duolingo.com/2021-duolingo-language-report/

This section has looked at multiple digital data sources and what they can bring to the field of migration studies. Clearly, digital trace data have huge potential given their timeliness and wide geographic availability. However, calibrating new data sources with and validating them against traditional data are essential to use novel sources effectively for migration analysis and policy. New digital data offer possibilities to study a diverse range of topics, including the scale of migration, intentions to migrate, and integration and cultural assimilation of migrants. Given their wide applicability to often politically sensitive topics, such as migration and human displacement, social scientists should critically reflect on the risks of results being misinterpreted, or, worse, misused, and how unethical uses of the data could harm individuals, particularly those in vulnerable situations, and infringe upon their fundamental rights (Beduschi, 2017). While many of the applications of computational social science to study are motivated by a potential positive impact on both migrants and the wider society, similar methods could be used to limit freedom and rights of migrants (for a comprehensive analysis of ethical considerations, see Taylor, 2023).

18.3 New Opportunities in Migration Research

The Digital Revolution has brought not only new data sources but also opportunities to apply new methodologies or augment research possibilities. Modelling migration is necessary because of the lack of quality in migration data from both traditional and digital sources. Digital trace data needs to be calibrated with traditional data. A natural way of combining data sources is through Bayesian models; indeed, Alexander et al. (2020) suggest a framework to combine migration data from multiple sources over time through a Bayesian hierarchical model. One level of the model focuses on adjusting the bias related to non-representative data (e.g. digital trace data) for a 'gold standard' given by survey data (e.g. the American Community Survey). Rampazzo et al. (2021) proposed a Bayesian hierarchical model as well. Their model combines traditional and digital data considering both data sources to be biased. Both frameworks stress that digital trace data cannot be a substitute for traditional data sources and that more accurate results can be obtained through their combination, rather than replacement.

Moreover, social media could also be actively used to recruit survey respondents. Advertisements on social media can be repurposed to recruit survey participants to answer a questionnaire. Facebook and Instagram have been used to recruit survey respondents during the COVID-19 pandemic (Grow et al., 2020), LGBTQ+ minorities (Kühne & Zindel, 2020), but also migrants (Pötzschke & Braun, 2017; Pötzschke & Weiß, 2021). Recruiting migrant respondents for traditional sampling strategies is notoriously challenging. However, social media advertising platforms such as that offered by Facebook provide the opportunity for non-probabilistic

sampling of migrants, through the use of the migration variable.[7] Pötzschke and Braun (2017) used Facebook to sample Polish migrants in four European countries—Austria, Ireland, Switzerland, and the UK. In the 4 weeks during which the ads were running, a total of 1100 respondents were recruited with a budget of 500 euro. Moreover, Pötzschke and Weiß (2021) used a similar design on Facebook and Instagram to recruit German migrants worldwide. 3800 individuals completed the questionnaire from 148 countries. The advantage of this strategy is to recruit migrant respondents worldwide in a timely manner and with modest budgets. However, it is challenging to produce representative results as there is no control over who opts in to the survey. This necessitates techniques such as post-stratification to make the results more representative of the specific migrant population. It may be worth noting that similar techniques are also used in traditional surveys (e.g. re-weighting, re-calibration), though with surveys on social media, the lack of a probability sampling results in a necessity to post-stratify.

Narratives around migration are usually investigated through qualitative interviews (Flores, 2017; Rowe et al., 2021). The proliferation of social media has also increased the volume of publicly available text that can be analysed to study general perceptions, narratives, and sentiments on a variety of topics. For instance, Twitter can also be used to analyse sentiments towards migrants and migration (Flores, 2017; Rowe et al., 2021). In 2010, the state of Arizona implemented an anti-immigrant law, the effect of which was studied using 250,000 tweets with natural language processing (NLP) techniques and a difference-in-difference design (Flores, 2017). Analysing the content of the tweets, the author stressed that policies have an effect on the perception of migrants, proving that micro-blogging data are an alternative source for public opinion on migrants (Flores, 2017). In Europe as well, analysis of Twitter text data delivered insights on sentiment towards migrants, describing a situation of polarisation of opinion (Rowe et al., 2021). The data provide an opportunity to track population sentiment towards migration in close to real time and monitor shifts over time. Moreover, focusing on the language used on social media, NLP might be useful to identify migrants and study migration flows (Kim et al., 2020).

High-intensity (e.g. weekly or monthly) time series are an opportunity to monitor change and create early alert systems for shifting migration patterns. Napierała et al. (2022) proposed a cumulative sum model to detect changes in trend of asylum applications. The use of flow data and early warning systems could help policymakers in anticipating refugee movements and improve preparedness and management capacities, if handled ethically and responsibly. However, these data and models can be used to make it more difficult for individuals to exercise their rights under the International Human Rights Law. Administrative data sources hold great potential for the study of migration patterns but present specific issues: for instance, their coverage is limited to the extent that people officially register or de-register from countries' administrative systems; also, administrative records track

[7] On Facebook Advertising Platform, it is possible to also create advertisements on Instagram.

events (e.g. asylum applications), not individuals, and are affected by issues of double-counting and biases that may affect their usability for official migration statistics. Eurostat data on number of applications lodged (which might also be biased) in EU countries could be augmented by including digital trace data in the model, increasing the ability to potentially anticipate future trends. This approach is suggested by Carammia et al. (2022) through an adaptive machine learning algorithm which combines data from Google Trends and traditional data sources. Given their frequency, data from social media platforms and Google Trends could indeed contribute to the early identification of shifting trends and, if managed responsibly, to greater capacities of migration policymakers and practitioners to inform adequate and timely measures (Alexander, Polimis and Zagheni, 2020; Martín et al., 2020).

Projects like Refugee.Ai and GeoMatch[8] propose to use data-driven algorithms to assign refugees across countries and improve their integration prospects (Bansak et al., 2018). Providing examples for the USA and Switzerland, Bansak et al. (2018) describe an algorithm based on supervised machine learning and optimal matching which takes into account the refugee characteristics (e.g. age, gender, language, education) and local site characteristics. The authors bring evidence of an improvement in subsequent refugee employment outcomes (from 34 to 48%). Moreover, they suggest that the model is flexible and can focus on different integration metrics to optimise for. The matching system is described also in the context of the UK (Jones & Teytelboym, 2018). Similar systems have been suggested also in Sweden to match refugees and property landlords (Andersson & Ehlers, 2020). Nevertheless, automated decisions should always be accompanied by a human element of review to avoid risks of algorithmic bias and human rights infringements.

There is evidence that also computational methods such as machine learning and neural networks might provide insights on migration. Simini et al. (2021) suggested a gravity model with deep neural network to predict flows of migrants and demonstrated that the model performed better than other models due to its geographic agnosticism. Moreover, convolutional neural networks might lead to new ways of fusing data and master high-frequency data (Pham et al., 2018).

18.4 The Way Forward

This chapter has demonstrated how the Digital Revolution has provided new data sources and opportunities to researchers. Timely data on migration are important not only for academics but also for policymakers and practitioners to design data-driven policies and programmes. The COVID-19 pandemic has stressed the importance of having timely and accurate mobility data for the study of the diffusion of the

[8] See https://immigrationlab.org/project/harnessing-big-data-to-improve-refugee-resettlement/.

virus (Alessandretti, 2022). However, data from digital traces often lacks a clear definition of what is being measured. Since such data are obtained from private companies, there may be no information available about the algorithms used to produce migration and mobility estimates, for example, about the specific criteria used to classify migrants. A clearer understanding of the construction of these measures would allow to include these data sources in models with more precision.

In the future, it would be important to create sustainable systems for safe and secure access to the data. At the moment, much of this research is dependent on application programming interfaces (API), which as attested by Freelon (2018) might be closed suddenly. When APIs are not available, web-scraping[9] might be a solution, but terms and conditions of the project as well as ethical implications should be taken into account. Initiatives such as the *Big Data for Migration Alliance* (BD4M),[10] convened by IOM's Global Migration Data Analysis Centre (GMDAC), the EU Commission Knowledge Centre on Migration and Demography (KCMD), and the Governance Lab (GovLab) at New York University, aim to provide a platform for cross-sectoral international dialogue and for guidance on ethical and responsible use of new data sources and methods. *Social Science One*[11] tries to create partnerships between academic researchers and businesses. At the moment, it has an active partnership with Facebook, established in April 2018. The initiative is led by Gary King (Harvard University) and Nathaniel Persily (Stanford University). The goal is to give researchers access to Facebook's micro-level data after having submitted a research proposal. There are significant privacy concerns from this, however, which has created delays in the process. On February 13, 2020, the first Facebook URLs dataset was made available; 'The dataset itself contains a total of more than 10 trillion numbers that summarize information about 38 million URLs shared more than 100 times publicly on Facebook (between 1/1/2017 and 7/31/2019)'.[12] A research proposal is needed to apply for access to such datasets; this is the first step in analysing large micro-level datasets from private social media companies. Companies also often control the analysis produced with their data. Researchers using companies' data have to follow strict contracts on its use and seek approval on the results before publication. The Social Science One initiative is interesting in this regard as it comes with pre-approval from Facebook. However, it also highlights challenges of relying on Facebook-internal teams to prepare the data in a non-transparent matter: recently, Facebook had to acknowledge that, accidentally, half of all of its US users were left out of the provided data.[13] This

[9] Web-scraping is defined as the process of automatically capturing online data from online websites (Marres & Weltevrede, 2013).

[10] https://data4migration.org

[11] https://socialscience.one

[12] https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one

[13] https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists/

essentially invalidated any work done with the data so far, including that of PhD students. To avoid such issues, ultimately caused by a lack of external oversight, researchers are increasingly calling for legally mandated corporate data-sharing programmes to enable outside, independent researchers to analyse and audit the platforms[14] (Guess et al., 2022).

Overall, the value of new data sources and new models cannot be underestimated. However, applications of these tools for research and public policy purposes should follow high ethical and data responsibility standards. New data sources and AI-based technologies could help researchers and policymakers improve prediction abilities and fill information gaps on migrants and migration, but the use of these technologies should be closely scrutinised and comprehensive risk assessments undertaken to ensure migrants' fundamental rights are safeguarded. The purposes of machine learning- and AI-based applications should be clearly communicated, and participatory approaches that empower migrant communities and 'data subjects' more generally should be promoted in research and policy domains, with a view to increasing transparency and public trust in these applications, but also provide guarantees for the protection of individual fundamental rights (Bircan & Korkmaz, 2021; Carammia et al., 2022). Many technologies come with a risk of being used to create 'digital fortresses'[15] in which these tools keep out migrants, rather than support them. Hence, social scientists and other researchers should carefully weigh the risks and potential repercussions when using digital traces.

References

Alessandretti, L. (2022) What human mobility data tell us about COVID-19 spread. *Nature Reviews Physics, 4*, 12–13.

Alexander, M., Polimis, K., & Zagheni, E. (2022). Combining social media and survey data to nowcast migrant stocks in the United States. *Population Research and Policy Review, 41*, 1–28. https://doi.org/10.1007/s11113-020-09599-3

Andersson, T., & Ehlers, L. (2020). Assigning refugees to landlords in Sweden: Efficient, stable, and maximum matchings. *The Scandinavian Journal of Economics, 122*, 937–965.

Avramescu, A., & Wiśniowski, A. (2021). Now-casting Romanian migration into the United Kingdom by using Google Search engine data. *Demographic Research, 45*, 1219–1254.

Azose, J. J., & Raftery, A. E. (2019) Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences, 116*, 116–122.

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*, 325–329.

[14] https://www.brookings.edu/research/how-to-fix-social-media-start-with-independent-research/

[15] https://apnews.com/article/middle-east-europe-migration-technology-health-c23251bec65ba45205a0851fab07e9b6

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*, 981–1018.

Bijak, J. (2010) *Forecasting international migration in Europe: A Bayesian view*. Springer Science & Business Media.

Billari, F. C. (2022). Demography: Fast and slow. *Population and Development Review, 48*, 9–30.

Billari, F. C., & Zagheni, E. (2017). Big data and population processes: A revolution. *Statistics and Data Science: New Challenges, New Generations*, In *Proceedings of the Conference of the Italian Statistical Society* (pp. 167–178). Firenze University Press, CC BY 4.0.

Bilsborrow, R. E., Hugo, G., Zlotnik, H., & Oberai, A. S. (1997). *International migration statistics: Guidelines for improving data collection systems*. International Labour Organization.

Bircan, T., & Korkmaz, E. E. (2021). Big data for whose sake? Governing migration through artificial intelligence. *Humanities and Social Sciences Communications, 8*, 1–5.

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*, 102347.

Bosco, C., Grubanov-Boskovic, S., Iacus, S., Minora, U., Sermi, F., & Spyratos, S. (2022). Data innovation in demography, migration and human mobility. arXiv preprint arXiv:2209.05460.

Boswell, C., Geddes, A., & Scholten, P. (2011) The role of narratives in migration policy-making: A research framework. *The British Journal of Politics and International Relations, 13*, 1–11.

Carammia, M., Iacus, S. M., & Wilkin, T. (2022). Forecasting asylum-related migration flows with machine learning and data at scale. *Scientific Reports, 12*, 1–16.

Carling, J., Erdal, M. B., & Talleraas, C. (2021) Living in two countries: Transnational living as an alternative to migration. *Population, Space and Place, 27*, e2471.

Cesare, N., Lee, H., McCormick, T., Spiro, E., & Zagheni, E. (2018) Promises and pitfalls of using digital traces for demographic research. *Demography, 55*, 1979–1999.

Cooksey, B. (2014). *An Introduction to APIs*. https://zapier.com/learn/apis/

Dubois, A., Zagheni, E., Garimella, K., & Weber, I. (2018) Studying migrant assimilation through facebook interests. In *International Conference on Social Informatics* (pp. 51–60). Cham: Springer.

Fiorio, L., Abel, G., Cai, J., Zagheni, E., Weber, I., & Vinué, G. (2017) Using twitter data to estimate the relationship between short-term mobility and long-term migration. In *Proceedings of the 2017 ACM on Web Science Conference - WebSci '17* (pp. 103–110). Troy, New York, USA: ACM Press.

Fiorio, L., Zagheni, E., Abel, G., Hill, J., Pestre, G., Letouzé, E., & Cai, J. (2021) Analyzing the effect of time in migration measurement using georeferenced digital trace data. *Demography, 58*, 51–74.

Flores, R. D. (2017). Do anti-immigrant laws shape public sentiment? A study of Arizona's SB 1070 using Twitter data. *American Journal of Sociology, 123*, 333–384.

Freelon, D. (2018) Computational research in the post-API age. *Political Communication, 35*, 665–668.

Gendronneau, C., Wiśniowski, 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: Exploring the potential of social-media data for measuring EU mobility flows and stocks of EU movers.* Publications Office of the European Union.

Grow, A., Perrotta, D., Fava, E. D., Cimentada, J., Rampazzo, F., Gil-Clavel, S., & Zagheni, E. (2020) Addressing public health emergencies via facebook surveys: Advantages, challenges, and practical considerations. *Technical Report*, SocArXiv.

Guess, A., Aslett, K., Tucker, J., Bonneau, R. and Nagler, J. (2021) Cracking open the news feed: Exploring what us Facebook users see and share with large-scale platform data. *Journal of Quantitative Description: Digital Media, 1*. https://doi.org/10.51685/jqd.2021.006.

Guidotti, R., Nanni, M., Giannotti, F., Pedreschi, D., Bertoli, S., Speciale, B., & Rapoport, H. (2020). Measuring immigrants adoption of natives shopping consumption with machine learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 369–385). Springer.

Halford, S., Weal, M., Tinati, R., Carr, L., & Pope, C. (2018). Understanding the production and circulation of social media data: Towards methodological principles and praxis. *New Media & Society, 20*, 3341–3358.

Herdağdelen, A., State, B., Adamic, L., & Mason, W. (2016). The social ties of immigrant communities in the United States. In *Proceedings of the 8th ACM Conference on Web Science*, WebSci '16 (pp. 78–84). New York, NY, USA: Association for Computing Machinery.

Jones, W., & Teytelboym, A. (2018). The local refugee match: Aligning refugees' preferences with the capacities and priorities of localities. *Journal of Refugee Studies, 31*, 152–178.

Kashyap, R. (2021). Has demography witnessed a data revolution? Promises and pitfalls of a changing data ecosystem. *Population Studies, 75*, 47–75.

Kennan, J., & Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica, 79*, 211–251.

Kim, J., Sîrbu, A., Giannotti, F., & Gabrielli, L. (2020) Digital footprints of international migration on twitter. In *International symposium on intelligent data analysis* (pp. 274–286). Springer.

Kühne, S., & Zindel, Z. (2020). Using facebook and instagram to recruit web survey participants: A step-by-step guide and application in Survey Methods: Insights from the Field (SMIF). Special issue: 'Advancements in Online and Mobile Survey Methods'. Retrieved from https://surveyinsights.org/?p=13558.

Kupiszewska, D., & Nowok, B. (2008). *Comparability of Statistics on International Migration Flows in the European Union* (pp. 41–71). Wiley.

Lee, E. S. (1966). A theory of migration. *Demography, 3*, 47–57.

Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013) Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday, 18*(5). https://doi.org/10.5210/fm.v18i5.4366.

Marres, N., & Weltevrede, E. (2013). Scraping the social? Issues in live social research. *Journal of Cultural Economy, 6*, 313–335.

Martín, Y., Cutter, S. L., Li, Z., Emrich, C. T., & Mitchell, J. T. (2020). Using geotagged tweets to track population movements to and from Puerto Rico after Hurricane Maria. *Population and Environment, 42*, 4–27.

Napierała, J., Hilton, J., Forster, J. J., Carammia, M., & Bijak, J. (2022). Toward an early warning system for monitoring asylum-related migration flows in Europe. *International Migration Review, 56*, 33–62.

Palotti, J., Adler, N., Morales-Guzman, A., Villaveces, J., Sekara, V., Herranz, M. G., Al-Asad, M., & Weber, I. (2020). Monitoring of the Venezuelan exodus through Facebook's advertising platform. *PLOS ONE, 15*, e0229175.

Petersen, W. (1958). A general typology of migration. *American Sociological Review, 23*, 256–266. http://www.jstor.org/stable/2089239.

Pham, K. H., Boy, J., & Luengo-Oroz, M. (2018). Data fusion to describe and quantify search and rescue operations in the Mediterranean sea. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 514–523). IEEE.

Pötzschke, S., & Braun, M. (2017). Migrant sampling using Facebook advertisements: A case study of Polish migrants in four European countries. *Social Science Computer Review, 35*, 633–653.

Pötzschke, S., & Weiß, B. (2021). Realizing a global survey of emigrants through Facebook and Instagram. https://doi.org/10.31219/osf.io/y36vr

Poulain, M., Perrin, N., & Singleton, A. (2006). *THESIM: Towards harmonised European statistics on international migration*. Presses universitaires de Louvain.

Rampazzo, F., Bijak, J., Vitali, A., Weber, I., & Zagheni, E. (2021). A framework for estimating migrant stocks using digital traces and survey data: An application in the united kingdom. *Demography, 58*, 2193–2218.

Rampazzo, F., & Weber, I. (2020). Facebook advertising data in Africa. *International Organization of Migration, Migration in West and North Africa and across the Mediterranean: Trends, Risks, Developments, Governance, 32*, 9.

Rowe, F., Mahony, M., Graells-Garrido, E., Rango, M., & Sievers, N. (2021). Using Twitter to track immigration sentiment during early stages of the COVID-19 pandemic. *Data & Policy, 3*, e36.

Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.

Simini, F., Barlacchi, G., Luca, M., & Pappalardo, L. (2021). A deep gravity model for mobility flows generation. *Nature Communications, 12*, 1–13.

Sîrbu, A., Andrienko, G., Andrienko, N., Boldrini, C., Conti, M., Giannotti, F., Guidotti, R., Bertoli, S., Kim, J., & Muntean, C. I. (2021). Human migration: The big data perspective. *International Journal of Data Science and Analytics, 11*, 341–360.

Sloan, L., & Quan-Haase, A. (2017) *The SAGE handbook of social media research methods*. SAGE.

Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2018). *Migration data using social media: A European perspective*. EUR 29273 EN.

Stewart, I., Flores, R., Riffe, T., Weber, I., & Zagheni, E. (2019). Rock, Rap, or Reggaeton?: Assessing Mexican Immigrants' Cultural Assimilation Using Facebook Data. *arXiv:1902.09453 [cs]*.

Stielike, L. (2022). Migration multiple? Big data, knowledge practices and the governability of migration. In *Research methodologies and ethical challenges in digital migration studies* (pp. 113–138). Cham: Palgrave Macmillan.

Sutherland, I. (1963). John Graunt: A tercentenary tribute. *Journal of the Royal Statistical Society: Series A (General), 126*, 537–556.

Taylor, L. (2023). Data justice, computational social science and policy. In *Handbook of computational social science for policy*. Springer.

UN (ed.) (1998). *Recommendations on statistics of international migration*. No. no. 58, rev. 1 in Statistical Papers. Series M. New York: United Nations.

US SEC Commision (2018). *Facebook Inc 2018 Annual Report 10-K*. https://www.sec.gov/Archives/edgar/data/1326801/000132680119000009/fb-12312018x10k.htm

US SEC Commision (2019). *Facebook Inc 2019 Annual Report 10-K*. https://sec.report/Document/0001326801-20-000013/fb-12312019x10k.htm

US SEC Commision (2020). *Facebook Inc 2020 Annual Report 10-K*. https://www.sec.gov/ix?doc=/Archives/edgar/data/1326801/000132680121000014/fb-20201231.htm

Wanner, P. (2021). How well can we estimate immigration trends using Google data? *Quality & Quantity, 55*, 1181–1202.

Willekens, F. (1994). Monitoring international migration flows in Europe: Towards a statistical data base combining data from different sources. *European Journal of Population, 10*, 1–42.

Willekens, F. (2019). Evidence-based monitoring of international migration flows in Europe. *Journal of Official Statistics, 35*, 231–277.

Zagheni, E., & Weber, I. (2015). Demographic research with non-representative internet data. *International Journal of Manpower, 36*, 13–25.

Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review, 43*, 721–734.

Zlotnik, H. (1987). The concept of international migration as reflected in data collection systems. *The International Migration Review, 21*, 925–946.