

Augmentation or replication? Assessing big data's role in migration studies

Tuba Bircan 

Department of Sociology, Vrije Universiteit Brussel, Brussels, Belgium
Email: tuba.bircan@vub.be

Received: 28 November 2023; **Revised:** 22 March 2024; **Accepted:** 04 June 2024

Keywords: big data; human migration; predictive models

Abstract

As the field of migration studies evolves in the digital age, big data analytics emerge as a potential game-changer, promising unprecedented granularity, timeliness, and dynamism in understanding migration patterns. However, the epistemic value added by this data explosion remains an open question. This paper critically appraises the claim, investigating the extent to which big data augments, rather than merely replicates, traditional data insights in migration studies. Through a rigorous literature review of empirical research, complemented by a conceptual analysis, we aim to map out the methodological shifts and intellectual advancements brought forth by big data. The potential scientific impact of this study extends into the heart of the discipline, providing critical illumination on the actual knowledge contribution of big data to migration studies. This, in turn, delivers a clarified roadmap for navigating the intersections of data science, migration research, and policymaking.

Policy Significance Statement

This research elucidates the transformative role of big data in migration studies, offering policymakers a compass for navigating the complexities of contemporary migration flows. It critically examines the integration of innovative data sources, such as mobile phone records and online activity, highlighting their potential to fill gaps in traditional migration statistics and enhance real-time responsiveness. The paper underscores the necessity for ethical stewardship and multidisciplinary collaboration in harnessing these data sources, ensuring that policies are not only informed by robust empirical evidence but are also aligned with ethical imperatives. By doing so, it provides a strategic blueprint for evidence-based policymaking that is attuned to the nuanced realities of migration and respects the dignity of individuals.

1. Introduction

The onset of the digital era has catalysed transformative changes across various academic disciplines, with the field of migration studies being a particularly notable example. The incorporation of big data analytics into this field is posited as a potential paradigm shift, heralding new levels of detail and immediacy that were once beyond the reach of traditional data sources such as censuses and surveys. Nonetheless, this technological progression prompts a critical inquiry: does big data merely reflect insights akin to

conventional methodologies, or does it genuinely enhance our comprehension of migration patterns with newfound information?

Historically, migration studies have predominantly relied on structured data sources which, despite their reliability, often trail in encapsulating the rapidly evolving dynamics of human mobility (Sirbu et al., 2021). In contrast, big data, with its myriad sources like internet-derived data (i.e., social media and online search trends), mobile phone records, satellite data, introduces a dynamic and instantaneous perspective of migration, potentially addressing voids left by traditional sources (Bircan and Korkmaz, 2021; Salah et al., 2022; Tjaden, 2021).

The exploration and integration of big data within migration research signify a paradigm shift that has been maturing over the past decade. Scholarly discourse increasingly recognised the transformative potential of big data to redefine the contours of migration studies. The introduction of big data into this field marks a significant departure from traditional methodologies towards more dynamic and real-time analytics. Pioneering works by Bengtsson et al. (2011), UN Global Pulse (2014), and Laczko and Rango (2014) have been instrumental in laying the foundation for applying big data analytics in migration research. Bengtsson et al. (2011) utilised mobile phone data after the Haiti earthquake to track migrations, setting a precedent for the application of big data in crisis situations. This approach further explored and demonstrated the advantages of real-time data from mobile phone records and social media over traditional sources in capturing migratory movements (Luca et al., 2021; Salah et al., 2019). Concurrently, UN Global Pulse (2014) investigated the potential of online search data to estimate migration flows, emphasising the significance of digital footprints in migration studies. Laczko and Rango (2014) critically assessed the “Migration Data Revolution,” advocating for the integration of big data to unravel complex migration patterns. This collective endeavour highlights a consensus on the need for innovative big data methodologies to achieve a comprehensive understanding of migration dynamics in the 21st century.

The European Commission’s Joint Research Centre (JRC) and the International Organization for Migration (IOM) have also been pivotal in incorporating big data into migration studies, significantly influencing theoretical and practical approaches. The JRC’s research, particularly highlighted by Gendronneau et al. (2019), has shown how big data can enhance our understanding of migration trends and inform policymaking, aiming to improve migrant conditions through data-driven insights. Similarly, the IOM’s World Migration Report series, notably discussed by McAuliffe and Martin (2021), has progressively embraced big data analytics, underscoring its vital role in global migration governance. These reports illustrate big data’s ability to offer detailed insights into migration patterns, facilitating more effective policymaking on a global scale.

The nuanced contributions of big data analytics to migration studies reveal a complex landscape where computational prowess and social science inquiry converge. This domain, enriched by technological innovation and empirical investigation, significantly deepens our understanding of migratory phenomena. The value of big data extends beyond the mere collection of large datasets; it fundamentally enhances our analytical capabilities concerning migration patterns.

However, while big data analytics offer invaluable new perspectives, they often necessitate corroboration with traditional data to maintain accuracy and representativeness (Spyratos et al., 2018, 2019). Consequently, big data should be viewed as a complementary tool that enriches, rather than supplants, traditional methodologies in migration studies (Bircan and Salah, *forthcoming*). Its real novelty lies in the methodological innovations it fosters, enabling researchers to identify patterns and predict trends previously unattainable. Nevertheless, it is crucial to recognise that despite these technological and methodological strides, progress in the governance, ethics and practice of big data in migration research has lagged. The ethical considerations and governance mechanisms necessary to navigate big data’s complex landscape remain underdeveloped, highlighting persistent challenges in maximising big data’s potential within ethical and governance frameworks (Beduschi, 2017). Therefore, the evolution of big data in migration studies serves as a vital reminder of the need for continued dialogue and development in these areas, ensuring its application promotes both research and the well-being of migrants in an ethically responsible manner.

The central inquiry of this paper is to discern the degree of innovation introduced by big data analytics in migration studies. This involves a comparative evaluation of both traditional and big data sources and their respective abilities to provide insightful, actionable and timely data for understanding and forecasting migration trends. The paper also seeks to examine how big data analytics can be effectively harnessed in policymaking, navigating through the challenges and opportunities in the data-driven governance of migration.

Additionally, this paper investigates the applications of big data sources in developing indicators for mobility and migration, scrutinising their added value in diverse areas like predicting asylum-seeker destination choices, understanding migrant integration, and reimagining migration policies. This exploration will provide a comprehensive evaluation of the current state and future potential of big data applications in migration research and policy development, emphasising the importance of a nuanced and informed utilisation of these emerging data sources (Salah et al., 2022; Tjaden, 2021). Having said that, this study does not intend to detract from the potential of big data in revolutionising migration studies. Rather, its objective is to critically assess its current utility and potential for offering innovative insights, especially in forecasting migration trends and shaping policy decisions. This encompasses an examination of the challenges entailed in integrating big data into established frameworks, such as concerns over data reliability, ethical considerations, and striking a balance between privacy and the utility of information (Salah, 2022; Sandberg et al., 2022).

In short, this study endeavours to make a significant contribution to the academic discourse on migration, offering a balanced and critical perspective on the role of big data in this dynamic field. In doing so, it seeks not only to advance our understanding of migration dynamics but also to furnish policymakers and practitioners with valuable insights for data-driven migration management and policy development.

2. Conceptual Approaches to Big Data for Migration

The incursion of big data into migration studies has precipitated not merely a technical revolution but a profound shift in the theoretical underpinnings of how migration is conceptualised and understood. Mayer-Schönberger and Cukier (2013) articulate this transformative shift by delineating big data's defining attributes—volume, variety, velocity, and veracity. These characteristics coalesce to afford a more intricate and nuanced analysis of migration, unearthing patterns, and trends that traditional datasets might not reveal, thus enriching the discourse on migration flows and their underlying determinants. Building upon this conceptual framework, Kitchin (2014) accentuates the transformative capability of big data to forge new insights via sophisticated analytical techniques such as machine learning (ML) and predictive analytics. These avant-garde methods are especially promising in refining the precision and temporal relevance of migration statistics and forecasts, presenting indispensable tools for researchers and policymakers (Bircan and Salah, *forthcoming*).

The socio-technical paradigm, as expounded by Bijker and Law (1992), furnishes an essential lens through which to examine the nexus between social contexts and technological infrastructures. This paradigm highlights the imperative to consider the interlaced social, ethical and political dimensions within data analytics. Leonelli (2015) propounds this argument further within the ambit of data science, arguing that data is a socio-technical construct, reflective of the values, biases and aims of those who generate and manipulate it. This perspective is acutely relevant in migration studies, where policies informed by data must be sensitive to the complex socio-political intricacies of migration and the rights and well-being of migrants.

Czaika and de Haas (2017) and Napierała et al. (2022) underscore the paradigmatic shift in migration policy formulation brought about by data science. The application of predictive analytics in particular offers a pathway towards more proactive, informed and nuanced migration policies. This is especially pertinent in scenarios that demand the anticipation of migration flows, such as those arising from climate change or geopolitical instability, where policies need to be both efficacious and compassionate. Furthermore, the intersection of big data, AI and migration studies not only opens avenues for improved

migration statistics and predictive models but also raises crucial ethical and policy-related considerations. The potential of big data and AI to inform and shape migration policies must be balanced with concerns related to data privacy, ethical usage and the potential for bias. This balance is essential in ensuring that data-driven approaches in migration studies are not only technically sound but also ethically responsible and socially equitable.

The burgeoning synergy of big data, artificial intelligence (AI) and migration studies promises substantial enhancements in the realm of migration statistics and predictive modelling. This nexus, however, does not come without its challenges—it heightens the imperative for critical ethical and policy-related considerations. The burgeoning capabilities of big data and AI to inform and reshape migration policies necessitate a careful and judicious balance against concerns surrounding data privacy, ethical use and the inherent risk of bias. The attainment of this equilibrium is crucial to ensure that data-driven methodologies within migration studies are not only technically robust but are also rooted in ethical soundness and social justice.

The schematic framework in Figure 1 encapsulates the vital theoretical constructs that govern this balance: knowledge creation, the socio-technical paradigm, and the intricate interplay between data science and policy. These foundational elements are instrumental in comprehending and harnessing the full potential of big data and AI within migration studies. They provide the pillars upon which the enhancement of migration statistics and predictive models are constructed, while simultaneously emphasising the importance of considering policy implications and ethical considerations within this rapidly evolving field.

As we forge ahead, the challenge lies in the robust theorisation of big data applications for migration. It is imperative that these applications are not dominated solely by computational and data science approaches but are also interwoven with socio-political theories. Such a theoretical grounding is

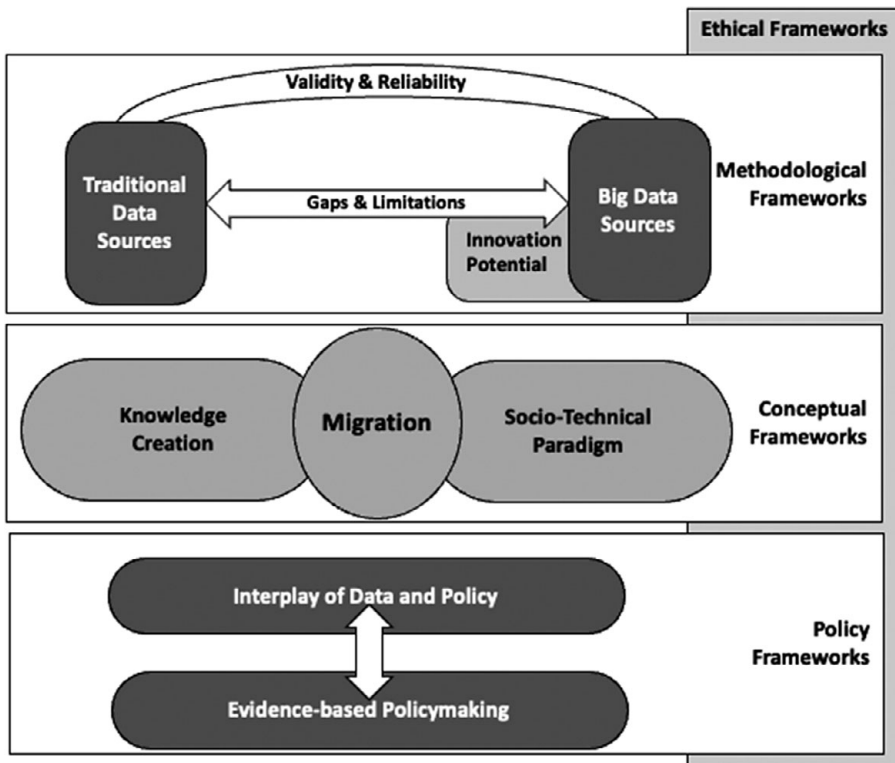


Figure 1. Conceptualization use of Big Data for measuring migration.

indispensable for informing policy implications and offering a navigational tool to steer through the complex landscape of migration governance.

Simultaneously, there is a critical need to refine traditional data sources, enhancing their quality, timeliness, geographical coverage, demographic breadth, and definitional clarity, as they provide the cornerstone for validating many big data-derived measurements. The improvement of traditional data is essential for better big data estimates, with big data applications demonstrating enhanced performance in areas such as internal migration and displacement. However, in regions where traditional data are sparse, big data sources also face limitations, often exacerbated by limited digital penetration in developing countries.

In response to this, ethical compliance, the involvement of a broader spectrum of social scientists, and strengthened collaboration across academia, the private sector and practitioners become paramount. These efforts are imperative to pave the way towards an effective, equitable and reliable utilisation of big data sources in migration studies. This collaborative and multidisciplinary approach forms the keystone for unleashing the potential of big data, crafting fair and actionable migration policies that are underpinned by ethical integrity and a deep respect for human rights. In essence, the fusion of these diverse approaches and disciplines will be pivotal in realising the vision of big data as a transformative force in the domain of migration policy and study.

3. What Do We Already Know? Assessing Big Data Sources for Migration Estimates

Migration studies have traditionally been informed by a suite of conventional data sources, each contributing unique insights into the multifaceted nature of human mobility. Population censuses, though robust, suffer from a lack of timeliness due to their decennial frequency, which significantly hampers their utility for capturing the dynamism of migration patterns (Laczko et al., 2023; Kraler and Reichel, 2022). Household surveys, such as the Demographic and Health Surveys (DHS) and the Living Standards Measurement Study (LSMS), offer granularity on migration-related attributes but are not without sampling biases, which may skew the representativeness of the data (McAuliffe and Triandafyllidou, 2021). Administrative records offer high accuracy for documented legal migrations but often fail to capture the undocumented or irregular movements, which are substantial elements of global migration (Kraler and Reichel, 2022). Vital statistics provide indirect indicators of migration through demographic changes, but their primary design is not for migration analysis, which limits their direct applicability (Reister et al., 2022).

In contrast, the past decade has seen a burgeoning interest in harnessing big data for migration estimates, with numerous studies exploring its potential to fill the gaps left by traditional methodologies (Bosco et al., 2022; Salah et al., 2022; Ahmad Yar and Bircan, 2023). The advent of big data has introduced a new epoch marked by digital footprints from mobile phone records, social media, and satellite imagery, which offer detailed, near real-time insights into migration flows and trends. These unconventional data streams have shown promise in improving migration statistics and developing predictive models, albeit with a progression that has been more measured than initially anticipated.

The integration of big data within migration studies, particularly through predictive models, represents a pivotal shift from traditional methodologies. The focus on predictive powers emerges from big data's ability to offer real-time, granular insights, surpassing the temporal and spatial constraints inherent in conventional data sources (Carammia et al., 2022; Anakal et al., 2024). Predictive modelling in migration studies is not merely about forecasting; it's a multifaceted approach that enhances understanding of migration patterns, informs policy decisions, and anticipates future trends. This emphasis on predictive analytics is driven by the need to address immediate and long-term migration challenges, utilising the wealth of information embedded within digital traces from mobile phone records, social media interactions, and satellite imagery. By developing predictive models, researchers can provide policymakers with tools to make informed decisions, adapt to changing migration dynamics, and implement proactive strategies. This shift towards predictive analytics signifies a broader trend in harnessing computational

techniques to decipher complex social phenomena, marking a significant evolution in the field of migration studies.

While big data sources have been leveraged to augment traditional migration statistics significantly, their application is not without challenges. Estimating migrant stocks has shown success when big data is combined with other data sources, yet the geographical coverage remains narrow. In developing countries, where data scarcity is most acute, big data's efficacy is hampered by limitations such as the limited use of the internet and social media (Sirbu et al., 2021; Kim et al., 2020). Furthermore, estimating cross-border mobilities and flows with big data is fraught with difficulties. Replicative studies have provided valuable insights, yet issues of data availability and comparability frequently emerge as stumbling blocks. Access to reliable data is a crucial determinant in these assessments, thus underscoring the necessity of data collaboratives for enhancing comparability and methodological consistency (Gendronneau et al., 2019; Rampazzo et al., 2023).

In Table 1, a systematic distillation of big data's contribution to migration studies is undertaken, delineating the development of specific migration indicators and their associated predictive capacities. *Mobile phone Call Detail Records* (CDRs) emerge as a pivotal source, facilitating the real-time tracking of internal displacement and migration patterns, with a particular novelty in their application to disaster response and the monitoring of disease spread. These CDRs are effectively utilised to parse internal migration patterns, with a nuanced application at the sub-regional level for international migration, drawing upon satellite data, census statistics, and social media to provide a holistic view of cross-border movements and integration processes. *Geo-located social media activity* from platforms such as Twitter and Facebook is leveraged to infer international migration flows and stocks with a granularity that allows for disaggregation by age, sex, skill levels or occupation. While the utility of this data in predictive models is not its primary focus, it nonetheless offers a dynamic, quasi-census that capitalises on user-generated data for demographic insights, with its efficacy enhanced when juxtaposed with official statistics for validation. *Online search data*, particularly from Google Trends, holds promise for estimating international mobility patterns and the forecasting of forced migration flows. This innovative approach is validated against traditional data sources, providing a forward-looking perspective on migration intentions and emergent flows, which has been recognised and employed by institutions such as the European Union Agency for Asylum (EUAA). *AI and ML techniques* stand at the frontier of predictive analytics in migration studies. They offer short-term predictive indices for migration flows, integrating variables such as market prices, rainfall, and conflict events. Moreover, ML techniques extend into the qualitative realm, analysing public attitudes towards refugees as discerned from radio content, requiring a synthesis of human expertise and varied data streams for a comprehensive analysis. The supplementary data sources serve a dual role—fortifying the data's credibility and providing additional dimensions to the indicators, thus enhancing the robustness of migration patterns analysis (Bosco et al., 2022; Carammia et al., 2022). The requirement for supplementary data, as highlighted for online search data like Google Trends, indeed extends across all big data sources. This necessity stems from the fact that big data, while rich and expansive, often lacks the contextual depth that traditional data provides. Whether it's CDRs, social media activity, or AI and ML outputs, each big data source benefits from being corroborated and enriched by additional data sets. These supplementary sources, such as official statistics, census data and qualitative assessments, provide the essential context, validation and verification that underpin the integrity and applicability of big data insights for migration studies. Thus, acknowledging the role of supplementary data across all big data sources is crucial in ensuring a robust and comprehensive approach to migration analysis.

To be more specific, Call Detail Records (CDRs) from mobile phones provide geo-located information that has proven invaluable in tracking population movements, especially in scenarios such as natural disasters or conflict-driven displacement. This area, enriched by the empirical investigations of scholars like Hughes et al. (2016) and furthered by Luca et al. (2021), underscores the utility of CDRs in analysing migration flows and enhancing national migration statistics, as demonstrated by Lai et al. (2019). Additionally, Pastor-Escuredo et al. (2019) highlight CDRs' value in estimating populations affected by displacement, aligning with the insights from Salah (2021) on humanitarian response challenges. Studies like Beine et al. (2021) and Tai et al. (2022) extend this narrative by employing CDRs to track

Table 1. Selection of Big Data applications for migration estimations and models

Big data source	Thematic areas of the developed migration indicators	Predictive models/powers	Supplementary sources needed	Novelty
Mobile Phone Call Detail Records (CDRs)	Tracking internal displacement and migration patterns	Used for internal migration patterns and international migration at the sub-regional level	Combined with satellite data for cross-border movements, census statistics for refugee integration, and social media data for social integration assessment	Novel use in disaster response and disease spread monitoring, provides real-time internal migration data (e.g., Blumenstock, 2012; Salah et al., 2019; Silm et al., 2020)
Geo-located Social Media Activity (Twitter, Facebook)	Inferring international migration flows and stocks, disaggregated by age, sex, skill levels, or occupation	Useful for understanding mobility patterns, but not specifically focused on predictive models	This could be enhanced with official statistics for validation and additional context	Offers a dynamic, real-time census at a national or global level, leverages user self-reported information for demographic insights (e.g., Sirbu et al., 2021; Goglia et al.; 2022; Hsiao et al., 2023)
Online Search Data (Google Trends)	Estimating international mobility patterns and predicting forced migration	Used to forecast migration, compare with numbers of asylum-seekers and migrants	Requires traditional data sources for context and verification	Innovative in predicting emigration flows and migration intentions based on search term analysis, used by the European Union Agency for Asylum (EUAA) for asylum application forecasts (e.g., Böhme et al., 2012; Bosco et al., 2022; Qi and Bircan, 2023)
Satellite Data (Remote Sensing)	Assessing the impact of natural disasters, environmental factors and land use changes on population movements	Data processing requires advanced computing, GIS analytics and remote sensing products are used for examining environment-mobility nexus such as displacement.	In need of ground truthing with official statistics, administrative data for validation and combinations with other (big) data sources	Can rapidly estimate displaced populations in post-disaster contexts, the impact of urbanisation and land use change (e.g., Checchi et al., 2013; Bircan, 2022; Kato and Lee, 2022)
AI and ML	Short-term predictions of migration flows, understanding public attitudes towards refugees	AI indices predict migration based on variables like market prices, rainfall and conflicts; ML analyses radio content for public attitudes	Enhanced by human expertise in interpreting big data from various sources for comprehensive analysis and predictive modelling	AI offers a sophisticated approach to predicting short-term migration trends, ML provides insights into public sentiments (i.e., Carrammia et al., 2020; Schutte et al., 202; Anakal et al., 2024)

short-term human mobility patterns, particularly useful in scenarios of internal displacement or emergency response, supported by Silm et al. (2020)'s findings on CDRs' resolution in tracking human movements. However, the ethical dimensions surrounding data privacy and the governance of CDR usage, as discussed by Salah et al. (2022), call for more stringent data governance frameworks to address these pressing concerns ensuring that the innovation in migration studies via big data aligns with ethical research practices and data protection standards.

One of the most striking examples of big data evolution is the use of Google search data to anticipate migration movements. The volume of Google searches has been shown to reflect the real-time public interest or concerns about migration-related issues, serving as an innovative proxy for migration intentions (Bosco et al., 2022). The pioneering work by Böhme et al. (2012) opened avenues for such analyses, demonstrating how digital footprints could offer insights into migration dynamics. Building on this, Qi and Bircan (2023) expanded the scope by utilising Google Trends data to predict forced migration patterns, an approach that has shown considerable promise in understanding and anticipating migration trends in response to crises. Further enriching this analytical repertoire, Avramescu and Wiśniowski (2021) applied these techniques to now-cast Romanian migration into the United Kingdom, while Carammia et al. (2020) used Google search data as a proxy for intention and combined with numerous data sources to predict asylum-related migration flows. Similarly, Wanner (2021) and Leysen and Verhaeghe (2023) have explored the efficacy of Google data in estimating immigration trends, including Japanese migration to Europe. Jurić (2022) extended this methodology to predict refugee flows from Ukraine, demonstrating big data's potential in crisis scenarios. Complementing these quantitative approaches, qualitative insights from Pew Research Center studies, such as Noe-Bustamante et al. (2020) and Connor (2020), provide a contextual understanding of migration discourses and policy impacts on movement. These diverse efforts showcase the evolving sophistication in utilising digital traces to inform migration policy.

Social media data has also emerged as a rich repository of user-generated content and metadata, offering new perspectives on migration sentiment and trends. Research by Goglia et al. (2022) and Hsiao et al. (2023) integrated social media data with official statistics to improve migration flow predictions. Despite its potential, this data source grapples with challenges concerning reliability and representativeness. Cairtin et al. (2021) explored these complexities, emphasising the need for robust methodologies to ensure the credibility of insights derived from social media analytics.

Furthermore, satellite imagery has been increasingly employed to monitor environmental conditions affecting migration, such as urbanisation patterns and responses to disasters. The work by Chen (2020) has shown, a focus on indirect indicators of human migration, such as changes in land use or vegetation, which can be precipitated by various factors including urbanisation and environmental changes. While this approach offers a broad-scale view of migration-related environmental changes, it also faces hurdles in ensuring accurate causal inference (Jain, 2020) and ethical compliance (Bircan, 2022), thereby requiring careful implementation and interpretation. The real-time, extensive coverage of satellite data provides a unique perspective often inaccessible through ground surveys. Checchi et al. (2013) have demonstrated its value in rapidly estimating displaced populations, essential for quick response initiatives. Furthermore, Camargo et al. (2020) unravel the interplay between migration, conflict and anthropogenic changes, offering insights crucial for policy and practice. Yet, to fully realise its potential, satellite imagery should be corroborated with ground-truth data to validate findings and add context, as Kato and Lee (2022) have done in their investigation of migration's impact on landslides. Satellite data's application in migration studies signifies notable progress but requires a comprehensive, ethical approach that Bircan (2022) insists must integrate various data sources, ensuring that our understanding of migration is not only precise but also ethically and empirically sound.

Advancing beyond the frontiers established by mobile phone data, social media analytics, internet search trends and satellite imagery, methodological advancement have been required to improve computing capacity. AI and ML as computational disciplines are the keystones in constructing and interpreting the sophisticated predictive models that now inform our understanding of migratory phenomena. Also, AI and ML emerge as pivotal instruments for computational migration research, refining our predictive

capabilities beyond the purview of traditional analytics. The ingenious application of these technologies by Carammia et al. (2020) and Anakal et al. (2024) exemplifies their transformative power, unveiling intricate migration patterns and providing foresight into asylum-related flows. ML algorithms excel in discerning the subtleties within vast datasets, identifying correlations and causations that inform not only the academic narrative but also the policy framework. In this intricate lattice of human mobility, AI dissects complex networks, extracting sentiments and attitudes with a deftness that rivals human intuition, as evidenced by the insightful work of Bansak et al. (2018). The scholarship in this domain does not merely exploit AI and ML as mere analytical tools but integrates them as central to a holistic approach to migration studies. This integration calls for a judicious synthesis of data integrity and computational innovation, ultimately driving forward the ethical application of big data in social science research.

This shift towards Big-Data-driven inquiry represents a pivotal move in modern migration studies, where Big Data is complemented with traditional datasets and other Big Data sources, enriching the empirical foundation from which policy decisions are sculpted. Yet, the guardianship of such data necessitates a judicious approach to managing the tangled balance between open access and the protection of individual data rights. Therefore, the ethical governance of Big Data in migration studies demands the establishment of stringent protocols and ethical guidelines to safeguard against the misuse of information and to assure that the quest for knowledge does not override the primacy of individual rights and ethical standards.

3.1. Ethics of Big Data and its relevance for migration policy

The integration of big data into migration studies and policymaking has thus far been a cautious journey, tempered by methodological, representational, and ethical considerations. While big data presents significant opportunities to address existing gaps in traditional migration statistics, its efficacy is still evolving (Bosco et al., 2022). Big data's role in migration policy is scrutinised through ethical lenses, considering the nuances of methodological and representational challenges. The current state of big data applications is predominantly at an assessment level, with a focus on nowcasting rather than forecasting. This approach is largely due to the challenges in validating new data sources, given the absence of high-quality, granular traditional data. While big data offers promising avenues for enhancing migration analysis, particularly in identifying gaps in migration statistics, the reliability and validity of these new measures are contingent upon overcoming the limitations of existing data frameworks. This juxtaposition highlights the critical need for high-quality, granular data to validate and harness big data's capabilities fully, moving beyond nowcasting to more predictive and comprehensive models (Spyratos et al., 2019). The ethical dimension remains paramount, ensuring that the application of big data in this field aligns with robust ethical standards, safeguarding the integrity and applicability of research outcomes for informed policymaking.

Addressing these challenges requires a multifaceted approach that does not solely rely on big data but employs it in tandem with traditional data sources. This hybrid methodology aims to capitalise on the real-time, large-scale capabilities of big data while anchoring its findings in the more established methods of traditional data analysis. However, this approach must navigate the inherent biases present in both data types. The effective integration of big data into migration studies is a nuanced and complex process, necessitating a careful examination of its limitations and potential biases to ensure a comprehensive and dependable understanding of migration patterns and public sentiments.

As we move forward, the need for more collaborative efforts between industry, academia and government becomes ever more apparent (Verhulst and Young, 2019, 2023). Such collaborations are vital for improving access to big data, enabling a more extensive analysis and robust policymaking. In this context, big data is not always as 'big' as required—not in terms of volume but in terms of availability and comparability. The current challenge is not only to continue the scientific exploration of big data sources for developing new indicators but also to focus on enhancing computational approaches to improve predictive models. Predictive models, which are inherently theory-driven, can only be validated through statistical assumptions that comply with traditional data. Therefore, while policymakers claim to rely on

evidence-based decision-making, there must be a prudent consideration of the quality, validity and reliability of the “evidence,” particularly for big data sources.

In summary, the potential of big data in migration studies is apparent, but its application must be informed by a comprehensive understanding that includes ethical considerations and practical utility. The advancement of the field will necessitate maintaining a critical perspective, continuously assessing the impact of methodological choices on the validity of research findings and their implications for policy. The collaborative efforts of researchers, practitioners, and policymakers will be paramount in realising the full potential of big data to enhance our understanding of migration and to craft policies that are both effective and respectful of human dignity and rights.

4. Migration, data, and policy nexus

The integration of big data into migration policy formulation marks a pivotal shift in the landscape of migration governance. The capacity to leverage vast, diverse data from sources such as mobile phone records, social media activity, satellite imagery, and internet search trends has equipped policymakers with more nuanced and dynamic tools for decision-making. This evolution is significantly acknowledged by international institutions like the OECD, IOM, EU, EC and the World Bank, which have shown a keen interest in employing innovative methodologies and big data analytics for measuring and estimating migration at both stock and flow levels as well as integration indicators (Bircan et al., 2023; IOM, 2021).

Big data analytics heralds a new era in migration studies, offering granular, timely and dynamic insights that traditional data sources struggle to match (Sirbu et al., 2021). These novel data sources offer a more immediate and nuanced perspective on migration flows, enabling the development of responsive and informed migration policies. For instance, the EC’s Knowledge Centre on Migration and Demography (KCMD) employs big data to shape migration policies, utilising social media data, mobile phone records and satellite imagery to gain insights into migration patterns (Gendronneau et al., 2019).

A compelling example of big data’s potential is its use following the 2010 Haiti earthquake, where mobile phone data was used to track population movements, aiding disaster response and recovery efforts (Bengtsson et al., 2011). Similarly, in Sweden, social media analytics have been used to gauge migration sentiments and trends, influencing policies related to integration and public perception (Salah et al., 2019).

The utilisation of predictive analytics derived from big data enables the anticipation of migration trends, a critical aspect for developing proactive and responsive migration policies (IOM, 2023b; Melachrinou et al., 2020). However, while big data applications excel in replicating known indicators and validating existing knowledge, their potential contributions extend beyond mere replication. The challenge lies in navigating ethical concerns and privacy issues, which, if not adequately addressed, may confine big data usage to an assessment level rather than a predictive or policy-shaping tool.

One promising area of development is the enhancement of models through computational approaches. These approaches are more theory-oriented and provide quantitative testing of theories, which is crucial for maintaining statistical assumptions and ensuring the robustness of predictive models. Policymaking, often touted as being evidence-based, must exercise caution regarding the quality, validity and reliability of “evidence,” especially when derived from big data sources. To support fair and actionable policy frameworks, it is imperative to link conceptual, methodological and ethical frameworks effectively. The following recommendations are posited for policymakers and practitioners in this evolving field:

- **Develop Transparent Data Governance Frameworks:** It is essential for policymakers to establish clear guidelines on the usage of big data in migration studies, ensuring adherence to international data protection regulations such as the EU GDPR (Data, E.G., 2018).
- **Foster Multi-Stakeholder Collaboration:** Collaboration with various stakeholders, including international organisations, NGOs, and the private sector, can enrich the quality and scope of big data analytics, as demonstrated by initiatives like the IOM’s Displacement Tracking Matrix (IOM, 2023a).

- **Invest in Capacity Building:** Building the capacity of policymakers and practitioners to interpret and utilise big data analytics is crucial for effective and informed migration governance, as suggested by entities like UN Global Pulse (Hidalgo-Sanchis, 2021).
- **Ensure Ethical Data Usage:** Upholding ethical practices in big data applications is paramount. This involves safeguarding the rights and privacy of individuals, as emphasised by researchers like Salah et al. (2022).
- **Embrace a Human-Centric Approach:** Decision-making processes should always prioritise the welfare of migrants and communities, ensuring that policies are grounded in a human-centric approach, akin to the UNHCR's Project Jetson (UNHCR, 2021).
- **Improve Data Accuracy and Representativeness:** Policymakers should focus on enhancing the precision and representativeness of big data, addressing potential biases to ensure reliable migration-related policymaking (Beduschi, 2021).
- **Promote Interdisciplinary Research:** Encouraging collaboration between data scientists and migration experts can foster innovative approaches and ensure that big data insights are contextually grounded and relevant, as discussed by Resnyansky (2019).
- **Advocate for Global Data Collaboratives:** Supporting international efforts to share migration-related data can enhance the global understanding of migration trends and foster cooperation in policy formulation (Salah, 2021; Verhulst 2021).

By embracing these recommendations, policymakers can effectively utilise the potential of big data to inform and enhance migration policies, while maintaining the highest standards of ethical practice and ensuring the dignity and rights of individuals are respected. This approach not only enhances the reliability and validity of migration policies but also ensures they are ethically sound and socially responsible.

5. Discussion: the future of big data in migration studies

This paper's journey through the evolving landscape of big data in migration studies concludes with an acknowledgment of a field on the brink of a significant transformation. The insights garnered from the various sections of this study underscore a methodological renaissance, one that is marked by both innovation and challenge. The integration of big data into migration research has brought about an analytical depth previously unattainable, with AI and ML techniques pushing the boundaries of predictive analytics and forecasting capabilities.

Our review has confirmed that, when applied judiciously, big data can greatly enhance traditional migration statistics, filling in gaps with its granularity and immediate applicability. The use of mobile phone records, social media analytics and online search data opens up new avenues for understanding intricate migration dynamics, particularly in the realms of internal migration and displacement. Nonetheless, these advancements are not without their limitations. The ethical quandaries and privacy issues that accompany big data's application in migration studies are significant, necessitating a framework for transparent, responsible data governance that does not infringe upon individual rights.

As we gaze into the future, the role of big data in migration studies promises to be even more revolutionary, with continuous methodological advancements expected to sharpen the precision of migration forecasts. The potential for big data to inform and guide real-time policy decisions paints a picture of a future where migration governance is both proactive and informed.

However, the fruition of this promise hinges on overcoming current limitations. The representativeness and biases present within big data, compounded by the digital divide—especially pronounced in developing regions—pose substantial challenges to the reliability and equity of big data-derived insights. Moreover, the reliance on computational methodologies necessitates a stronger incorporation of socio-political theoretical frameworks to ensure that these approaches remain grounded in the complex realities of human migration. Furthermore, this discussion brings to the fore the indispensable role of traditional data sources. Improvements in the quality, timeliness, geographical coverage and demographic details of

traditional data are crucial, as they provide the grassroots validation for many big data-derived measurements. Better traditional data sources are a prerequisite for the enhancement of big data estimates, reinforcing the need for a symbiotic relationship between the two.

In light of these discussions, the path forward must involve a holistic approach that integrates ethical compliance and fosters collaboration across academia, the private sector and practitioners. The engagement of more social scientists in big data projects will be instrumental in addressing the ethical and theoretical challenges present in current methodologies.

In closing, the intersection of big data and migration studies is laden with potential that must be navigated with care and responsibility. The advancement of the field will require a persistent critical perspective, continually assessing the impact of methodological choices on research validity and policy implications. It is through the collective efforts of researchers, practitioners, and policymakers that we can harness the full potential of big data, ensuring that our enhanced understanding of migration leads to the formulation of policies that are not only effective but also respectful of human dignity and rights.

Author contribution. Conceptualization: T.B; Data curation: T.B; Formal analysis: T.B; Investigation: T.B; Methodology: T.B; Writing, review and editing: T.B.

Funding statement. This research is supported by the European Commission through the Horizon2020 European project: “HumMingBird – Enhanced migration measures from a multidimensional perspective” (GA: 870661). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interest. None declared.

Provenance. This article is part of the Data for Policy 2024 Proceedings and was accepted in Data and Policy on the strength of the Conference’s review process.

References

- Ahmad Yar AW and Bircan T (2023). Big data for official migration statistics: Evidence from 29 national statistical institutions. *Big Data and Society*, 10(2), 20539517231210244.
- Anakal S, Ravish GK, Sowjanya MN, Thejaswini MN, Mahalakshmi M and Gaikwad SM 2024. Predictive analytics and machine learning in assessing migration patterns: A comparative study, *Migration Letters*, 21(S4), pp. 1557–1564.
- Ashford LS, Kaneda T and Letouzé E 2022. Demystifying big data for demography and global health, *Population Bulletin*, 76(1).
- Avramescu A and Wiśniowski A 2021. Now-casting Romanian migration into the United Kingdom by using Google Search engine data, *Demographic Research*, 45, pp. 1219–1254.
- Beine M, Noy I, and Parsons C (2021). Climate change, migration and voice. *Climatic Change*, 167(1-2), 8.
- Beduschi A 2017. The big data of international migration: Opportunities and challenges for states under international human rights law, *Geo. J. Int’l L.*, 49, pp. 981–1002.
- Bengtsson L, Lu X, Thorson A, Garfield R and Von Schreeb J 2011. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti, *PLoS Medicine*, 8(8), e1001083.
- Bijker WE and Law J (Eds.). (1994). *Shaping technology/building society: Studies in sociotechnical change*. Cambridge, MA: MIT Press.
- Bircan T 2022. Remote sensing data for migration research, In Salah, A.A., Korkmaz, E.E., and Bircan, T. (Eds.) *Data Science for Migration and Mobility Studies*, Oxford University Press, pp.121–148.
- Bircan T and Korkmaz EE (2021). Big data for whose sake? Governing migration through artificial intelligence. *Humanities and Social Sciences Communications*, 8(1), 1–5.
- Bircan T and Salah AA (Forthcoming). Emerging Big Data Sources for Studying Migration. In Vargas-Silva C, Markaki Y and Allen W (Eds.), *Handbook of Research Methods in Migration* (2nd edition). Northampton, MA: Edward Elgar.
- Bircan T, Salah AA, and Sirbu A (2023). How Can Big Data Analytics Help Understand Migrant Integration? In *Harnessing Data Innovation for Migration Policy: A Handbook for Practitioners* (pp. 92–103). Geneva: International Organization for Migration.
- Blumenstock JE (2012). Inferring patterns of internal migration from mobile phone call records: Evidence from Rwanda. *Information Technology for Development*, 18(2), pp. 107–125.
- Bosco C, Grubanov-Boskovic S, Iacus S, Minora U, Sermi F and Spyratos S (2022). Data Innovation in Demography, Migration and Human Mobility, EUR 30907 EN, Publications Office of the European Union, Luxembourg.
- Böhme MH, Gröger A, and Stöhr T (2020). Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, 142, 102347.
- Caitrin A, Ate P, Zook M, Derek R, and Soehl T (2021). Challenges when identifying migration from geo-located Twitter data. *EPJ Data Science*, 10(1).

- Camargo G, Sampayo AM, Peña Galindo A, Escobedo, FJ, Carriazo, F, and Feged-Rivadeneira, A.** (2020). Exploring the dynamics of migration, armed conflict, urbanization, and anthropogenic change in Colombia. *Plos One*, 15(11), e0242266.
- Carammia M, Iacus S and Wilkin T** (2020). Forecasting asylum-related migration flows with machine learning and data at scale, *Scientific Reports*, 12, 1457.
- Cecchi F, Stewart BT, Palmer JJ, and Grundy C** (2013). Validity and feasibility of a satellite imagery-based method for rapid estimation of displaced populations, *International Journal of Health Geographics*, 12, 1-12.
- Chen X** (2020). Nighttime lights and population migration: Revisiting classic demographic perspectives with an analysis of recent European data. *Remote Sensing*, 12(1), 169.
- Connor P** (2020). More than nine-in-ten people worldwide live in countries with travel restrictions amid COVID-19. Available at: <https://policycommons.net/artifacts/616270/more-than-nine-in-ten-people-worldwide-live-in-countries-with-travel-restrictions-amid-covid-19/1596889/>. [Accessed 08 Mar 2024].
- Czaika M and de Haas H** (2017). The effect of visas on migration processes. *International Migration Review*, 51(4), 893–926.
- Data EG** (2018). EU General Data Protection Regulation (GDPR). Available at: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>; last accessed 12 November 2023.
- Gendronneau C, Wiśniowski A, Yildiz D, Zagheni E, Fiorio L, Hsiao Y and Hoorens S** (2019). *Measuring Labour Mobility and Migration Using Big Data*. Brussels: European Commission.
- Goglia D, Pollacci L and Sirbu A** (2023). Dataset of Multi-Aspect Integrated Migration Indicators. *Data*, 8(9), 139.
- Hidalgo-Sanchis P** (2021). UN Global pulse: a UN innovation initiative with a multiplier effect. In Lapucci M., and Cattuto C., (eds) *Data Science for Social Good: Philanthropy and Social Impact in a Complex World* (pp. 29–40). Cham: Springer.
- Hsiao Y, Fiorio L, Wakefield J and Zagheni E** (2023). Modeling the Bias of Digital Data: An Approach to Combining Digital with Official Statistics to Estimate and Predict Migration Trends. *Sociological Methods and Research*, 00491241221140144.
- Hughes C, Zagheni E, Abel GJ, Sorichetta A, Wisniowski A, Weber I and Tatem AJ** (2016). Inferring migrations: traditional methods and new approaches based on mobile phone, social media, and other big data: feasibility study on inferring (labour) mobility and migration in the European union from big data and social media data. Available at: <https://ec.europa.eu/jrc/>; [Accessed 12 November 2023].
- International Organization for Migration (IOM)**. (2023a). Displacement Tracking Matrix. Geneva: IOM.
- International Organization for Migration (IOM)**. (2023b). *Harnessing Data Innovation for Migration Policy: A Handbook for Practitioners*. Geneva: IOM.
- Jurić T** (2022). Predicting refugee flows from Ukraine with an approach to Big (Crisis) Data: a new opportunity for refugee and humanitarian studies, *Medrxiv*, 2022-03.
- Kashyap R** (2021). Has demography witnessed a data revolution? Promises and pitfalls of a changing data ecosystem, *Population Studies*, 75(sup1), 47–75.
- Kato Y, and Lee J** (2022). Quantitative analysis of the effect of human migration on the landslides after the 2010 Haiti earthquake using Landsat imagery, *Natural Hazards*, 111(3), 3031–3052.
- Kim J, Sirbu A, Giannotti F and Gabrielli L** (2020, April). Digital footprints of international migration on twitter. In *International Symposium on Intelligent Data Analysis* (pp. 274–286). Cham: Springer International Publishing.
- Kitchin R** (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society*, 1(1), 2053951714528481.
- Kraler A and Reichel D** (2022). Migration Statistics. In Scholten, p. (Ed) *Introduction to Migration Studies An Interactive Guide to the Literatures on Migration and Diversity* (pp. 439). Springer Nature.
- Laczko F and Rango M** (2014). Can big data help us achieve a “Migration Data Revolution”? *Migration Policy Practice (IOM)*, 4 (2), pp. 20–29.
- Laczko F, Vidal EM, and Rango M** (2023). *Measuring Global Migration: Towards Better Data for All*. Taylor and Francis.
- Lai S, Erbach-Schoenberg EZ, Pezzulo C, Ruktanonchai NW, Sorichetta A, Steele J and Tatem AJ** (2019). Exploring the use of mobile phone data for national migration statistics. *Palgrave Communications*, 5, 1–10.
- Leonelli S** (2015). What counts as scientific data? A relational framework. *Philosophy of Science*, 82(5), 810–821.
- Leyser B and Verhaeghe PP** (2023). Searching for migration: estimating Japanese migration to Europe with Google Trends data, *Quality and Quantity*, 57(5), 4603–4631.
- Luca M, Barlacchi G, Oliver N, and Lepri B** (2021). Leveraging mobile phone data for migration flows. *arXiv preprint arXiv: 2105.14956*.
- Mayer-Schönberger V and Cukier K** (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston, MA: Houghton Mifflin Harcourt.
- McAuliffe M and Triandafyllidou A** (2021). *World Migration Report 2022*. Geneva: International Organization for Migration (IOM).
- Melachrinou C, Carammia M and Wilkin T** (2020). Using big data to estimate migration “push factors” from Africa. In *Migration in West and North Africa and across the Mediterranean: Trends, risks, development and governance* (pp. 98–116).
- Napierala J, Hilton J, Forster JJ, Carammia M and Bijak J** (2022). Toward an early warning system for monitoring asylum-related migration flows in Europe. *International Migration Review*, 56(1), 33–62.
- Noe-Bustamante L, Mora L and Lopez MH** (2020). About one-in-four US Hispanics have heard of Latinx, but just 3% use it. *Pew Research Center*, 11.
- Qi H and Bircan T** (2023). Can Google Trends predict asylum-seekers’ destination choices? *EPJ Data Science*, 12(1), 41.

- Pastor-Escuredo D, Imai A, Luengo-Oroz M, and Macguire D** (2019). Call detail records to obtain estimates of forcibly displaced populations. In ‘*Guide to Mobile Data Analytics in Refugee Scenarios: The ‘Data for Refugees Challenge’ Study*’, pp. 29–52.
- Rampazzo F, Rango M, and Weber I** (2023). New Migration Data: Challenges and Opportunities. In Bertoni E, Fontana M, Gabrielli L, Signorelli S and Vespe M (Eds) *Handbook of Computational Social Science for Policy* (pp. 345). City, State: Publisher Name.
- Recchi E, and Tittel K** (2023). The Empirical study of human mobility: Potentials and pitfalls of using traditional and digital data. In Bertoni E, Fontana M, Gabrielli L, Signorelli S and Vespe M (Eds) ‘*Handbook of Computational Social Science for Policy*’. Cham: Springer International Publishing, pp. 437–464.
- Reister M, MacFeely S, Me A, Hereward M, Schweinfest S, and Warschburger S** (2022). A statistical quality assurance framework for the United Nations. *Statistical Journal of the International Association of Official Statistics*, 38, 725–732.
- Resnyansky L** (2019). Conceptual frameworks for social and cultural Big Data analytics: Answering the epistemological challenge, *Big Data and Society*, 6(1), 2053951718823815.
- Salah AA** (2021). Mobile data challenges for human mobility analysis and humanitarian response. In ‘*Research Handbook on International Migration and Digital Technology*’, pp. 106–122.
- Salah AA** (2022). Can big data deliver its promises in migration research? *International Migration*, 60(2), 252–255.
- Salah AA, Bircan T and Korkmaz EE** (2022a). New data sources and computational approaches on migration and human mobility. In Salah AA, Korkmaz EE and Bircan T (Eds.) *Data Science for Migration and Mobility Studies* (pp. 3–23). Oxford: Oxford University Press.
- Salah AA, Canca C and Erman B** (2022b) Ethical and legal concerns on data science for large scale human mobility. In Salah AA, Korkmaz EE and Bircan T (Eds.) *Data Science for Migration and Mobility. Proceedings of the British Academy*. London: Oxford University Press.
- Salah AA, Pentland A, Lepri B, Letouzé E, De Montjoye YA, Dong X and Vinck P** (2019). *Guide to Mobile Data Analytics in Refugee Scenarios. The ‘Data for Refugees Challenge’ study*. Cham: Springer.
- Sandberg M, Rossi L, Galis V and Bak Jørgensen M** (2022). *Research Methodologies and Ethical Challenges in Digital Migration Studies: Caring For (Big) Data?* Cham: Springer Nature.
- Schutte S, Vestby J, Carling J and Buhaug H** (2021). ‘Climatic conditions are weak predictors of asylum migration’, *Nature communications*, 12(1), 2067.
- Silm S, Järv O and Masso A** (2020). Tracing human mobilities through mobile phones. In *Handbook of Research Methods and Applications for Mobilities* (pp. 182).
- Sirbu A, Andrienko G, Andrienko N, Boldrini C, Conti M, Giannotti F and Sharma R** (2021). Human migration: The big data perspective. *International Journal of Data Science and Analytics*, 11, 341–360.
- Spyratos S, Vespe M, Natale F, Weber I, Zagheni E and Rango M** (2018). *Migration data using social media*. JRC Science Hub.
- Spyratos S, Vespe M, Natale F, Weber I, Zagheni E and Rango M** (2019). Quantifying international human mobility patterns using Facebook Network data. *PLoS One*, 14(10), e0224134.
- Tai XH, Mehra S and Blumenstock JE** (2022). Mobile phone data reveal the effects of violence on internal displacement in Afghanistan. *Nature Human Behaviour*, 6(5), 624–634.
- Tjaden J** (2021). Measuring migration 2.0: a review of digital data sources, *Comparative Migration Studies*, 9(1), 59.
- UN Global Pulse**. (2014). Estimating Migration Flows Using Online Search Data. *Global Pulse Project Series*, 4, pp. 1–2.
- UNHCR**. (2021). Jetson project. Available at: <https://jetson.unhcr.org> [Accessed 12 November 2023].
- Verhulst SG** (2021). ‘Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs’, *Data and Policy*, 3, e6.
- Verhulst SG, and Young A** (2019). The potential and practice of data collaboratives for migration. *Guide to Mobile Data Analytics in Refugee Scenarios: The ‘Data for Refugees Challenge’ Study*, 465–476.
- Verhulst S, and Young A** (2023). Operationalizing Data Collaboratives for Migration. *Harnessing Data Innovation for Migration Policy: A Handbook for Practitioners* (pp. 92–103). Geneva: International Organization for Migration.
- Wanner P**. (2021). How well can we estimate immigration trends using Google data?, *Quality and Quantity*, 55(4), 1181–1202.