

2023, Vol. 4(3), 795-807
© The Author(s) 2023
Article reuse guidelines:
<https://dergi.bilgi.edu.tr/index.php/reflektif>
DOI: 10.47613/reflektif.2023.139
Article type: Research Article

Received: 27.09.2023
Accepted: 03.10.2023
Published Online: 09.10.2023

Sedef Turper*

Big Data in Migration Research: The Ethical Considerations Göç Çalışmalarında Büyük Veri: Etik Değerlendirmeler

Abstract

The current study initiates a discussion on the ethical dilemmas inherent in big data research within the domain of migration studies. It encourages a conversation regarding the guiding principles that researchers should uphold for the responsible utilization of data. As such, the study focuses on the challenges migration researchers face in designing, conducting, and disseminating their research findings. Accordingly, this study raises considerations related to the privacy of both individuals and groups, as well as to the creation and exacerbation of inequalities and power asymmetries. It also emphasizes the crucial role that migration researchers can play in establishing ethical standards for big data research.

795

Öz

Bu çalışma, göç çalışmaları alanında gerçekleştirilen büyük veri araştırmalarında karşılaşılan etik zorluklara odaklanmakta ve verinin sorumlu bir şekilde kullanımı için izlenmesi gereken ilkeler hakkında bir diyaloga katkı sunmayı amaçlamaktadır. Çalışmada göç araştırmacılarının çalışmalarını tasarlama, yürütme ve bulgularını yayma süreçlerinde karşılaştıkları zorluklara odaklanılmaktadır. Bu bağlamda, çalışmada araştırmacıların birey ve grup gizliliği ile güç dengesizliklerinin arttırımı ve yeni eşitsizliklerin yaratılmasına ilişkin etik sorumlulukları olmalıdır. Çalışmada ayrıca göç araştırmacılarının büyük veri araştırmaları için etik standartların belirlenmesindeki kritik rolü vurgulanmaktadır.

Keywords

Migration studies, big data, research ethics, group privacy, power asymmetry

Anahtar Kelimeler

Göç çalışmaları, büyük veri, araştırma etiği, grup gizliliği, güç asimetrisi

* Koç University, sturper@ku.edu.tr, ORCID: 0000-0001-5659-3181.

Introduction

The advent of social big data, generated through automated logs, has unlocked a level of precision and granularity previously beyond the reach of researchers (Vinck et al., 2019). Its integration into migration studies stands as a seminal point, fundamentally enhancing our comprehension of migration dynamics. Particularly, the wealth of communication data sourced from mobile phones and online platforms has proven invaluable, enabling us to map migration flows and dissect the intricate process of integration (Sirbu et al., 2021). However, the impact of this data revolution extends far beyond the augmentation of research tools; it reshapes the very practice of research, prompting profound reflections on the ethical dimensions of big data research within the field of migration studies. The real-time or near-real-time features, coupled with the predictive capacities of AI models, present a formidable ethical challenge, demanding careful consideration (Mahoney et al., 2022; Zimmer, 2018). Given the increasing use of pioneering Big Data and AI applications in the field of migration research and management, and considering the vulnerability of populations involved, migration scholars are uniquely positioned to explore the ethical challenges.

796

This study embarks on a critical exploration of the ethical dilemmas intrinsic to big data research in the domain of migration studies. My focus centers on the responsible use of Big Data in forced migration research, scrutinizing the attendant risks tied to individual and group privacy, as well as the potential for exacerbating power imbalances and creating new disparities.

The next section provides a concise overview of big data and its applications within the domains of humanitarian and developmental studies, underscoring its pivotal role in shaping migration scholarship. Subsequently, an examination of the ethical challenges arising in the context of big data research in migration studies is introduced. This discussion commences by addressing ethical considerations pertaining not only to individual privacy but also to privacy of the groups. Subsequently, it delves into the complex terrain of adhering to the ‘do no harm’ principle when disseminating and utilizing research findings by underscoring the risks of potentially creating or exacerbating inequalities and power asymmetries. The paper closes with concluding remarks that highlight the pivotal role that migration researchers would play in shaping ethical standards of big data research.

What is Big Data and What Venues Does It Offer for Migration Research?

Big data refers to exceptionally large and complex datasets that exceed the capabilities of traditional data processing methods, demanding advanced computational techniques for storage, analysis, and visualization (Chen et al., 2014). In comparison to traditional datasets, big data typically is generated continuously (Kitchin, 2014) and characterized by its volume, velocity,

veracity, variety, and value (Naeem et al., 2022). Communication data stands out as a significant subset of big data, encompassing a wide array of information generated through various communication channels, including text messages, emails, social media interactions, phone call records, and more. As providing data and metadata increasingly became a requirement for individuals to gain access to online social networks and telecommunication services, private companies regularly generate data at a precision level that was formerly available only to intelligence agencies through automated logs of social activity (Vinck et al., 2019). This type of data holds valuable insights into social interactions, information flow, and patterns of human communication (Boyd & Crawford, 2012; Ruths & Pfeffer, 2014).

While policymakers increasingly turn to novel data sources and analytical strategies for identifying the target populations for efficient policy interventions (Stielike, 2022; Taylor, 2016), a growing number of scholars acknowledge big data as a transformative opportunity for opening new venues for scientific inquiry (van Dijck, 2014; Fuchs, 2017). Consequently, big data has recently emerged as a notable tool for humanitarian and developmental interventions across various domains. Previous research leveraged real-time or near-real-time information for enhancing responses to natural disasters using diverse data sources such as satellite imagery and communication data for the rapid identification of affected areas and damage assessment (Bagrow et al., 2011; Bengtsson et al., 2011; Tomaszewski, 2014). Big data has also been proven instrumental in tracking the propagation of infectious diseases (Peak et al., 2018; Wesolowski et al., 2012) and in offering insights into economic disparities (Decuyper et al., 2014; Jean et al., 2016; Pokhriyal & Jacques, 2017). Furthermore, big data analytics significantly contributed to informed decision-making in infrastructural development investments through the analysis of mobility patterns and urban data (Martinez-Cesena et al., 2015).

In recent years, the integration of big data in migration research has significantly enriched the understanding of migration dynamics (Abel & Sander, 2014). Social big data, particularly communication data, derived from mobile phones and online platforms, has proven invaluable in mapping migration flows, and dissecting integration outcomes. With its real-time or near-real-time capabilities, big data offers a level of granularity and immediacy that is often challenging to capture using conventional surveys or administrative records alone (Luca et al., 2021). The analytical potential of big data empowers researchers to scrutinize migration flows and stocks, investigate migrant integration, and glean insights into the reception of migrants within host communities, as well as their social ties.

Earlier research has shown that communication data is an exceptionally efficient way of tracing and predicting human mobility (de Montjoye et al., 2013). Mobile call data records (CDR) and geo-located time-stamped data have frequently been utilized for understanding human mobility. Initiatives like the Data for Development (D4D) challenge in Senegal and the Data for Refugees (D4R) challenge in Turkey motivated researchers to empirically analyze and theoretically model human mobility by utilizing communication data (Blondel et al., 2012;

Salah et al., 2018). Previous research employing social big data has significantly contributed to the understanding of individual and group mobility (Giannotti et al., 2011; González et al., 2009; Lulli et al., 2017; Pappalardo et al., 2015) as well as to the forecasting of mobility intentions (Böhme et al., 2020). Communication data is proven to be instrumental in delineating social interaction patterns among immigrants and discerning geo-spatial patterns of segregation (Gao et al., 2021). Moreover, online social network data and sentiment analysis have frequently been leveraged to capture sentiments towards immigrants and fluctuations in migration flows (Chen et al., 2022; Rowe et al., 2021).

Balancing Potential Benefits and Ethical Imperatives: Navigating Big Data in Migration Research

Big data presents valuable venues for scholars and practitioners in the domain of humanitarian and developmental research. Yet, paradoxically, the very attributes that render this data so precious also serve to amplify the risks to privacy, fairness, equality, and due process when employed for research and decision-making (Sandberg & Rossi, 2022). It is in this delicate balance between potential benefits and ethical considerations that the pressing need for an encompassing ethical framework arises. Such a framework is essential to guide the responsible use of big data, ensuring that it serves the broader public good while safeguarding the rights and dignities of the research subjects.

798

Recognizing the immense potential of big data as a catalyst for achieving the Sustainable Development Goals, the United Nations Global Pulse initiative, inaugurated in 2015, took swift action to expedite the development of regulatory frameworks (UN Global Pulse & GSMA, 2017; UN Global Pulse & IAPP, 2017). This pivotal step paved the way for a diverse cohort of experts to come together under the auspices of international humanitarian organizations and address the ethical challenges posed by the utilization of big data in humanitarian and developmental research. Their collaborative efforts yielded a set of ethical principles that extend beyond mere privacy norms, aiming to provide a compass for the conscientious application of big data in these crucial fields by also addressing the risks to fairness, equality, and due process (OCHA, 2016; The Humanitarian Data Science and Ethics Group, 2020; United Nations Development Group, 2017). While these guidelines offer invaluable guidance, their implementation requires careful consideration. It is incumbent upon researchers to interpret these guiding principles with care, ensuring that the potential benefits of the research far outweigh any potential risks or harms that may arise.

Navigating the ethical terrain of big data research is inherently challenging, a task that is further compounded when dealing with sensitive data on vulnerable populations. In research focused on migration, particularly forced migration, where highly sensitive data is utilized,

the dynamic interplay between vulnerability and sensitivity necessitates an increased level of precaution (Duvell et al., 2010). The confluence of forced migration research and data science introduces a heightened level of ethical complexity, amplifying key considerations in migration research. While issues pertaining to privacy, confidentiality, security, and autonomy have always been pivotal in any migration study (Krause, 2017; Zapata-Barrero & Yalaz, 2020), additional concerns emerge at the nexus of forced migration and data science. This intersection demands researchers to exercise even greater vigilance in safeguarding the welfare and rights of their research subjects, given the sensitive nature of the data involved and the vulnerability of the populations under study.

The advent of digital migration studies has enabled researchers to analyze migration exclusively through data, bypassing direct engagement with the migrants who generate this information. This shift towards a more media-centric approach in ethical considerations, prioritizing the data over the individuals behind it, underscores the need for a more expansive and nuanced ethical framework in the realm of digital migration studies (Sandberg & Rossi, 2022). Scholars are increasingly recognizing the importance of adopting a more carefully expanded ethical approach in this evolving field. However, while migration scholars advocate for a paradigm shift beyond the minimal standard of ‘doing no harm,’ emphasizing the recognition and promotion of migrants’ agency and autonomy (Mackenzie et al., 2007), even upholding this basic standard can be a formidable task for researchers engaged in big data research, particularly within the domain of forced migration. The most challenging aspect of adhering to the ‘do no harm’ principle in digital migration studies lies in the conceptualization of vulnerability as a forward-looking variable (Thywissen, 2006), encompassing a predictive dimension of potential harm. This necessitates researchers to conduct a comprehensive assessment of both immediate and future risks, taking into account the socio-political, economic, cultural, and technological contexts in which research subjects and populations are embedded.

The following section initiates a discussion on the ethical challenges that migration researchers encounter when designing, conducting, and disseminating their research findings. Rather than offering an exhaustive list of issues regarding the ethical implications of big data research in the field of migration studies, the discussion emphasizes considerations pertaining to the privacy of both individuals and groups, as well as the creation and exacerbation of inequalities and power asymmetries.

Considerations on and Beyond Individual Privacy

Ethical guidelines and data protection regulations pertaining to big data research place a distinct emphasis on the responsibilities of data holders and users in safeguarding data privacy. These guidelines primarily focus on concerns related to identifiable natural persons, advo-

cating for practices such as pseudonymization and anonymization of data (European Union Agency for Fundamental Rights, 2018). Additionally, principles like privacy-by-design methods, purpose limitation, and data minimalization are recommended to prevent the reidentification of anonymized personal data.

The issue of re-identification of anonymized data holds particular significance in migration research, as the identification of research subjects carries a high potential for harm. This risk is exacerbated in the case of forced migrants, as the re-identification of their data could have dire consequences, potentially exposing them to human traffickers or the authorities of the states they fled from (Duvell et al., 2008). Especially, real-time or near-time mapping of individual mobility introduces a further layer of risk of being re-identified (Pastor-Escuredo et al., 2019), potentially exposing undocumented migrants to persecution, detention, or deportation. This risk is particularly acute if their mobility is deemed to violate the laws of the host country or if they become entangled with human smuggling networks. In light of these considerations, researchers bear the ethical obligation to prioritize the privacy of individuals within their research. This responsibility encompasses not only structuring and storing data in a manner that safeguards individuals' privacy but also making critical decisions about which findings to disclose. This decision-making process must account not only for the current state of technology but also anticipate potential advancements that could jeopardize privacy.

However, while discussions on data privacy have predominantly focused on individual privacy, the concept of group privacy emerges as an equally pertinent concern for migration researchers. The advent of big data analytics introduces new dimensions to privacy considerations at the group level (Kammourieh et al., 2017). While individuals may find a degree of anonymity within a crowd, the crowd itself remains discernible. This challenges the prevailing notions of privacy and data protection, especially when the objective is not individual identification but the location of a specific group of interest (Taylor, 2016). This issue becomes particularly critical for displaced populations in contexts of ongoing conflicts. Compromising the group privacy of these populations in situations of civil conflict can have severe repercussions, potentially exposing individuals to persecution. Disseminating information on mobility patterns or residential information of displaced communities seeking refuge can lead to dire consequences for the individuals involved. Revealing information about the established routines of migrant populations also entails inherent risks, as it may make them vulnerable during periods of social unrest, as escalating tensions between host and migrant communities can quickly lead to potential outbreaks of violence against migrants. In the light of those considerations, researchers should develop a concept of privacy that extends beyond individual considerations and encompasses group privacy and other emerging forms of identifiability resulting from datafication. This expanded perspective acknowledges that privacy is not solely related to individuals, but also is a concern for the collectivity itself, particularly for communities that have

been uprooted. By recognizing and ensuring group privacy, digital migration researchers can contribute to the well-being and the protection of their research subjects' fundamental rights.

Creation or Exacerbation the Power Asymmetries and Inequalities

Mapping and predicting mobility patterns serve as invaluable tools for delivering essential aid to vulnerable communities, particularly in the wake of large-scale disruptions to social order brought about by natural or man-made crises. The development of real-time or near-real-time mapping of individual or group mobility holds immense potential for humanitarian efforts, enabling the identification of critical hotspots where interventions are most urgently needed. The effective projecting and monitoring hotspots in time and space proves instrumental in monitoring diverse needs spanning from shelter requirements to essential healthcare services of displaced populations including refugee communities (Pastor-Escuredo et al., 2019). However, it is imperative to acknowledge that the use of incomplete or inaccurate data in humanitarian research can have far-reaching consequences for the provision of services to affected populations. Flawed data can yield inaccurate planning of humanitarian and developmental interventions, potentially depriving individuals of crucial opportunities and essential services. Hence, researchers engaged in humanitarian research must maintain a keen awareness of the biases inherent in the data sources they utilize. This vigilance is crucial in order to prevent the inadvertent creation of disparities in the distribution of services and benefits among targeted communities.

Furthermore, the exacerbation of power asymmetries between individuals and states emerges as a central ethical challenge for researchers engaged in big data research within the field of migration studies. In a global landscape where high-income country governments possess a vested interest in controlling mobility, mainly through forecasting, tracking, and preventing unauthorized migration flows towards their borders, the ability to identify mobility patterns of individuals and groups becomes highly valuable (Taylor, 2016). The capability to monitor and forecast human mobility holds profound implications for border control and migration management, granting authorities the capacity to intervene prior to individuals attaining refugee or asylum-seeker status, potentially reshaping their rights to movement. Consequently, ethical considerations regarding the well-being of research subjects in migration studies bear significant resemblance to those in the field of criminology. Identification of individuals not adhering to mobility restrictions within the country of residence or entry may lead to the apprehension of research subjects (Duvell et al., 2010). This is particularly pertinent in research on forced migration, where subjects' movements are likely to contravene regulations governing international border crossings. Even in cases where migrants have obtained legal protection status, their movements may still pique the interest of law enforcement agencies, tasked with overseeing compliance with mobility restrictions within the host country.

In this context, the potential utilization of digital data, mapping technologies and predictive tools for dataveillance purposes, which involves the continuous monitoring of (meta)data for unstated preset purposes, in migration management raises substantial concerns (Bircan & Korkmaz, 2021; Latonero & Kift, 2018; Sandberg & Rossi, 2022). Technologies initially conceived for humanitarian purposes are prone to adoption for migration management. Collaborative ventures among governments, academia, and data firms, culminating in the co-creation of these innovative technologies and their exchange, further accelerates their commercial and administrative applications, intended to serve the interests of research participants (van Dijck, 2014). However, such technologies possess the potential to facilitate more conservative and preemptive migration policies, potentially encroaching upon the fundamental human rights of research populations (Molnar, 2019). Consequently, researchers find themselves at a pivotal juncture, tasked with discerning whether to disseminate research findings that may pique the interest of law enforcement agencies. This encompasses pivotal decisions regarding which agencies to collaborate with in developing technologies and methodologies that may also serve dataveillance purposes.

Researchers, therefore, need to contemplate the potential (mis)use of data and innovative technologies that hold the potential to breach the privacy not only of individuals but also of groups. In the context of big data research on forced migration, the existing power asymmetries between migrants and states necessitates the thoughtful positioning of researchers. They must evaluate whether sharing their findings or tools may directly advantage enforcement agencies, potentially exacerbating the existing imbalance at the expense of migrants' rights and well-being. Ultimately, researchers would need to engage in a comprehensive assessment of the potential risks and benefits of their work by conceptualizing risk as a forward-looking notion. In doing so, researchers can make ethically informed decisions regarding the dissemination of their findings and the formation of research collaborations that ensure no harm befalls the vulnerable research populations.

Conclusion

The current study embarked on a critical examination of the ethical challenges entailed in big data research within the field of migration studies. It aimed to stimulate a dialogue concerning the fundamental principles that researchers must uphold in the conscientious use of data. Central to this exploration were the multifaceted challenges confronted by migration scholars throughout the stages of designing, conducting, and disseminating their research endeavors.

Commencing with the considerations surrounding the privacy of research subjects, the paper delved into the responsibilities of safeguarding not only the privacy of individuals but also that of groups. In doing so, emphasis was placed on the similarities drawn between mig-

ration and criminology studies, and the potential risks associated with the re-identification of individuals, as well as with the dissemination of findings that may lead to the identification, apprehension, detention, or deportation of individuals who fail to comply with the mobility regulations of their respective countries. Additionally, it was highlighted that disseminating information on mobility patterns or residential information of displaced and migrant communities can have severe repercussions for the groups, potentially exposing them to persecution in contexts of ongoing conflicts and emerging social tensions.

The study further explored the intricate landscape of adhering to the 'do no harm' principle when disseminating and utilizing research findings, with a focus on the potential risks of creating or exacerbating inequalities and power asymmetries. Examination was conducted into the adverse consequences stemming from the use of incomplete or inaccurate data in humanitarian research in the provision of services, along with the likely use of data and innovative technologies for monitoring individuals for migration management purposes.

Ultimately, the study invites researchers to embark on a comprehensive assessment of the potential risks and benefits inherent in their work, grounding this assessment in a forward-looking conception of risk. The responsibility of migration scholars encompasses not only structuring and storing data in a manner that safeguards individuals' privacy but also recognition of group privacy to ensure their research subjects' well-being and the protection of their fundamental rights. Vigilant risk assessment by migration researchers also proves indispensable in preventing inadvertent disparities and avoiding the exacerbation of existing power imbalances between states and migrants, all while preserving the rights and well-being of the latter.

The digitalized migration studies, like other big data research, contributes to the legitimization of personal data usage, leading to greater trust in and public acceptance of the commercial and administrative applications of these technologies as well. Scholars engaged in Big Data research should thus work towards building the ethical standards that will guide the use of such technologies across all sectors, ensuring ethical parity between scholarly and commercial or administrative applications of such technologies. Given that Big Data and AI applications are frequently tested and trained on migrant populations (Molnar, 2019), scholars specializing in forced migration are uniquely positioned to provide invaluable insights. They wield the potential to uphold a critical perspective on the applications of these technologies in the field of migration, thereby safeguarding migrant and refugee communities against the potential overreach of governments and private entities.

References

- Abel, G. J., & Sander, N. (2014). Quantifying global international migration flows. *Science*, *343*(1520–1522), 4793–4795.
- Bagrow, J. P., Wang, D., & Barabási, A.-L. (2011). Collective response of human populations to large-scale emergencies. *PLoS ONE*, *6*(3), e17680.
- Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & 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. In *PLoS Medicine* (Vol. 8, Issue 8).
- Bircan, T., & Korkmaz, E. E. (2021). Big data for whose sake? Governing migration through artificial intelligence. *Humanities and Social Sciences Communications*, *8*(1), 1–5.
- Blondel, V. D., Esch, M., Chan, C., Clérot, F., Deville, P., Huens, E., Morlot, F., Smoreda, Z., & Ziemlicki, C. (2012). Data for development: The [D4D] challenge on mobile phone data. *CoRR*, *abs/1210.0*. <http://arxiv.org/abs/1210.0137>
- Böhme, M. H., Gröger, A., & Stühr, T. (2020). Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, *142*, 102347.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, *15*(5), 662–679.
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Network Applications*, *19*, 171–209.
- Chen, Y., Sack, H., & Alam, M. (2022). Analyzing social media for measuring public attitudes toward controversies and their driving factors: A case study of migration. *Social Network Analysis and Mining*, *12*(1).
- de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, *3*(1), 1376.
- Decuyper, A., Rutherford, A., Wadhwa, A., Bauer, J.-M., Krings, G., Gutierrez, T., Blondel, V. D., & Luengo-Oroz, M. A. (2014). *Estimating food consumption and poverty indices with mobile phone data*. *arXiv preprint arXiv:1412.2595*.
- Duvell, F., Triandafyllidou, A., & Vollmer, B. (2008). Ethical issues in irregular migration research. *CLANDESTINO Working Paper*.
- Duvell, F., Triandafyllidou, A., & Vollmer, B. (2010). Ethical issues in irregular migration research in Europe. *Population, Space and Place*, *16*(3), 227–239.
- European Union Agency for Fundamental Rights. (2018). Handbook on European data protection law 2018. In *Luxembourg: Publications Office of the European Union*. Publications Office of the European Union.
- Fuchs, C. (2017). From digital positivism and administrative big data analytics. *European Journal of Communication*, *32*(1), 37–49.
- Gao, Q. L., Yue, Y., Tu, W., Cao, J., & Li, Q. Q. (2021). Segregation or integration? Exploring activity disparities between migrants and settled urban residents using human mobility data. *Transactions in GIS*, *25*(6), 2791–2820.

- Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S., & Trasarti, R. (2011). Unveiling the complexity of human mobility by querying and mining massive trajectory data. *VLDB Journal*, 20(5), 695–719.
- González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2009). Understanding individual human mobility patterns. *Nature*, 458(7235), 238–238.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790–794.
- Kammourieh, L., Baar, T., Berens, J., Letouzé, E., Manske, J., Palmer, J., Sangokoya, D., & Vinck, P. (2017). Group privacy in the age of big data. In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), *Group Privacy: New Challenges of Data Technologies* (pp. 37–66). Springer.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society*, 1(1).
- Krause, U. (2017). Researching forced migration: Critical reflections on research ethics during fieldwork. *Refugee Studies Centre*
- Latonero, M., & Kift, P. (2018). On digital passages and borders: Refugees and the new infrastructure for movement and control. *Social Media + Society*, 4(1), 1–11.
- Luca, M., Barlacchi, G., Oliver, N., & Lepri, B. (2021). Leveraging mobile phone data for migration flows. *CoRR*, abs/2105.1. <https://arxiv.org/abs/2105.14956>
- Lulli, A., Gabrielli, L., Dazzi, P., Dell’Amico, M., Michiardi, P., Nanni, M., & Ricci, L. (2017). Scalable and flexible clustering solutions for mobile phone-based population indicators. *International Journal of Data Science and Analytics*, 4(4), 285–299.
- Mackenzie, C., McDowell, C., & Pittaway, E. (2007). Beyond “do no harm”: The challenge of constructing ethical relationships in refugee research. *Journal of Refugee Studies*, 20(2), 299–319.
- Mahoney, J., Le Louvier, K., Lawson, S., Bertel, D., & Ambrosetti, E. (2022). Ethical considerations in social media analytics in the context of migration: Lessons learned from a Horizon 2020 project. *Research Ethics*, 18(3), 226–240.
- Martinez-Cesena, E. A., Mancarella, P., Ndiaye, M., & Schläpfer, M. (2015). *Using Mobile Phone Data for Electricity Infrastructure Planning*. Retrieved from <http://arxiv.org/abs/1504.03899>
- Molnar, P. (2019). Technology on the margins: AI and global migration management from a human rights perspective. *Cambridge International Law Journal*, 8(2), 305–330.
- Naeem, M., Jamal, T., Diaz-Martinez, J., Butt, S. A., Montesano, N., Tariq, M. I., De-la-Hoz-Franco, E., & De-La-Hoz-Valdiris, E. (2022). Trends and future perspective challenges in big data. In J.-S. Pan, V. E. Balas, & C.-M. Chen (Eds.), *Advances in Intelligent Data Analysis and Applications* (pp. 309–325). Springer Singapore.
- OCHA, U. (2016). Building data responsibility into humanitarian action. In *OCHA Policy and Studies Series* (Issue May). Retrieved from <https://www.unocha.org/publication/policy-briefs-studies/building-data-responsibility-humanitarian-action>
- Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A. L. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications*, 6.

- Pastor-Escuredo, D., Imai, A., Luengo-Oroz, M., & Macquiere, D. (2019). Call detail records to obtain estimates of forcibly displaced populations. In A. A. Salah, A. Pentland, B. Lepri, & E. Letouzé (Eds.), *Guide to Mobile Data Analytics in Refugee Scenarios* (pp. 29–52). Springer.
- Peak, C. M., Wesolowski, A., Erbach-Schoenberg, E. zu, Tatem, A. J., Wetter, E., Lu, X., Power, D., Weidman-Grunewald, E., Ramos, S., Moritz, S., Buckee, C. O., & Bengtsson, L. (2018). Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: Use of mobile phone data. *International Journal of Epidemiology*, 47(5), 1562–1570.
- Pokhriyal, N., & Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 114(46), E9783–E9792.
- 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 and Policy*, 3(12).
- Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, 346(6213), 1063–1064.
- Salah, A. A., Pentland, A., Lepri, B., Letouzé, E., Vinck, P., de Montjoye, Y.-A., Dong, X., & Dağdelen, Ö. (2018). *Data for refugees: The D4R challenge on mobility of Syrian refugees in Turkey* (arXiv preprint arXiv:180700523).
- Sandberg, M., & Rossi, L. (2022). Caring for (big) data: An introduction to research methodologies and ethical challenges in digital migration studies. In M. Sandberg, L. Rossi, V. Galis, & M. B. Jørgensen (Eds.), *Research Methodologies and Ethical Challenges in Digital Migration Studies: Caring for (Big) Data?* (pp. 1–21). Palgrave Macmillan.
- Sîrbu, A., Andrienko, G., Andrienko, N., Boldrini, C., Conti, M., Giannotti, F., Guidotti, R., Bertoli, S., Kim, J., Muntean, C. I., Pappalardo, L., Passarella, A., Pedreschi, D., Pollacci, L., Pratesi, F., & Sharma, R. (2021). Human migration: The big data perspective. *International Journal of Data Science and Analytics*, 11(4), 341–360.
- Stielike, L. (2022). MigrationMultiple? Big data, knowledge practices and the governability of migration. In *Research Methodologies and Ethical Challenges in Digital Migration Studies: Caring for (Big) Data?* (pp. 113–138). Palgrave Macmillan.
- Taylor, L. (2016). Safety in numbers? Group privacy and big data analytics in the developing world. In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), *Group Privacy: New Challenges of Data Technologies* (pp. 13–36). Springer. <https://doi.org/10.2966/scip.140117.131>
- The Humanitarian Data Science and Ethics Group. (2020). *A framework for the ethical use of advanced data science methods in the humanitarian sector*. [https://www.migrationdataportal.org/sites/g/files/tmzbdl251/files/2020-06/Framework Advanced Data Science in The Humanitarian Sector.pdf](https://www.migrationdataportal.org/sites/g/files/tmzbdl251/files/2020-06/Framework%20Advanced%20Data%20Science%20in%20The%20Humanitarian%20Sector.pdf)
- Thywissen, K. (2006). *Components of risk - A comparative glossary*. United Nations University – Institute for Environment and Human Security.
- Tomaszewski, B. (2014). Geographic information systems (GIS) for disaster management. In *Geographic Information Systems (GIS) for Disaster Management*. CRC Press.

- UN Global Pulse, & GSMA. (2017). *The state of mobile data for social good report* (Issue June). Retrieved from https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/06/Mobile-Data-for-Social-Good-Report_29June.pdf
- UN Global Pulse, & IAPP. (2017). *Building ethics into privacy frameworks for big data and AI*. Retrieved from <https://www.unglobalpulse.org/2018/10/building-ethics-into-privacy-frameworks-for-big-data-and-ai-a-report-from-un-global-pulse-and-iapp/>
- United Nations Development Group. (2017). *UNDG guidance note on big data for achievement of the 2030 agenda : Data privacy, ethics, and protection*. 16. Retrieved from <https://undg.org/wp-content/uploads/2017/03/UNDG-Big-Data-Guidance-Note.pdf>
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance and Society*, 12(2), 197–208.
- Vinck, P., Pham, P. N., & Salah, A. A. (2019). “Do no harm” in the age of big data: Data, ethics, and the refugees. In A. A. Salah, A. Pentland, B. Lepri, & E. Letouzé (Eds.), *Guide to Mobile Data Analytics in Refugee Scenarios* (pp. 87–99). Springer.
- Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., & Buckee, C. O. (2012). Quantifying the impact of human mobility on malaria. *Science*, 338(6104), 267–270.
- Zapata-Barrero, R., & Yalaz, E. (2020). Qualitative migration research ethics: A roadmap for migration scholars. *Qualitative Research Journal*, 20(3), 269–279.
- Zimmer, M. (2018). Addressing conceptual gaps in big data research ethics: An application of contextual integrity. *Social Media + Society*, 4(2), 2056305118768300.